An Improved Measure Of Reading Skill:

The Cognitive Structure Test

Robert Sorrells & Bruce Britton

The University of Georgia

Final Report
Account # 10-21-RR274-106
US Navy

ROBERT CURTIS SORRELLS
An Improved Measure of Reading Skill: The Cognitive Structure Test
(Under the direction of BRUCE K. BRITTON)

This study compared the construct validity and the predictive validity of a new test, called the Cognitive Structure Test, to multiple-choice tests of reading skill, namely the Armed Forces Vocational Aptitude Battery Paragraph Comprehension Test (ASVABpc) and the Scholastic Aptitude Test (SAT). To test the hypothesis that the Cognitive Structure Test is a better test of reading skill, 347 Air Force recruits read 16 technical and scientific passages (8 from the ASVABpc and 8 from the SAT) and took both multiple-choice and Cognitive Structure tests on the passages. For each passage the recruits' Cognitive Structure Test responses were compared to the responses of two experts who also read the passage and took the Cognitive Structure Tests (a total of 32 experts). These experts also took the multiple-choice tests.

Results indicated that the two tests measured very similar constructs and had similar reliability (Chronbach's Alpha) ratings (Cognitive Structure Test, .63; Multiple-choice test, .68).

Results showed that the Cognitive Structure Test was superior to the multiple-choice tests in predicting final Technical School grade point averages for the recruits, and equivalent to the multiple-choice tests in predicting the ASVAB vocational and general science tests, and general knowledge tests. For the ASVAB passages, the Cognitive

Structure Test predicted the final Techschool scores, but the multiple-choice test did not. Both tests predicted domain specific knowledge and general ability.

Correlations between the Cognitive Structure responses of the two experts from each passage and the scores of the experts on the multiple-choice questions were reported. The mean correlation between Expert 1 and Expert 2's cognitive structure responses for the ASVABpc passages was .75, and for the SAT passages was .62. The experts correctly answered 93.8% of the ASVABpc passage multiple-choice questions, but only 82.9% of the SAT passage multiple-choice questions.

Graphical representations of the structures elicited by the Cognitive Structure Test were presented for two passages using Addtree (Sattath & Tversky, 1977) and Extree (Corter & Tversky, 1986).

Suggestions for improving the Cognitive Structure Test were made, along with a review of the unresolved issues.

INDEX WORDS: Cognitive Structures, Reading Skill

REPORT DOCUMENTATION PAGE

FORM APPROVED
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1997	3. REPORT TYPE AND DATES COVERED Final Report

4. TITLE AND SUBTITLE OF REPORT An improved measure of reading skill: Cognitive Structures	5. FUNDING NUMBERS G10-21-RR274-106
6. AUTHOR(S) Robert Sorrells & Bruce K. Britton	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Dr. Robert C. Sorrells & Dr. Bruce K. Britton Department of Psychology University of Georgia, Athens	8. PERFORMING ORGANIZATION REPORT NUMBER:
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) ONR	10. SPONSORING/MONITORING AGENCY REPORT NUMBER:

11. SUPPLEMENTARY NOTES:

12a. DISTRIBUTION AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This study compared the validity of the Cognitive Structure Test (CST) to the Armed Forces Vocational Aptitude Battery Paragraph Comprehension Test (ASVAB) and the Scholastic Aptitude Test. 347 Air force recruits read 16 tecnical passages and took the CST and the ASVAB and SAT tests on the passages, and were compared to 32 experts who also read the passages and took the tests.
Results indicated that the tests were very similar in reliability and validity. The CST was superior to the other tests in predicting final Technical School grades. The tests were equivalent in measuring other general and specific knowledge.
Suggestions for the use and improvement of the CST were made, along with a review of the unresolved issues.

14. SUBJECT TERMS Reading skill. Cognitive Structure Testing.	15. NUMBER OF PAGES: 128
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT: unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT unclassified	20. LIMITATION OF ABSTRACT

TABLE OF CONTENTS

Introduction

Just and Carpenter (1987) have argued that standard reading tests were constructed atheoretically with no detailed model of how a reader analyzes the text's organization. The focus of test development, they contend, has been largely statistical, relying on norm referenced or criterion referenced approaches to establishing predictive validity. In fact, the predictive validity of these tests was quite good (Lavin, 1965). However, little concern was given to the psychological processes that underlie the reader's comprehension of a text that subsequently produced the given score. Predictive validity took precedence over construct validity. Yet psychometric theory gives priority to construct validity, especially in the behavioral sciences (Nunnally, 1978).

The construct validity of multiple-choice tests has been questioned for many years. Many studies have demonstrated that multiple-choice reading comprehension questions can often be answered without the presence of the passage (Pyrczak, 1972; Tuinmann, 1973). In a series of studies, Katz and his colleges (Katz, Lautenschlager, Blackburn, & Harris, 1990; Katz, Blackburn, & Lautenschlager, 1991; Katz & Lautenschlager, 1994) have demonstrated that

this problem extends to widely used college entrance exams (e.g., SAT, GRE, ACT). The results of these studies demonstrated that students could correctly answer a substantial percentage of reading comprehension questions when the passages accompanying the questions were not presented. These questions were presumably designed to measure the ability of readers to obtain information and draw inferences from text. If readers can answer these questions above chance levels without ever reading the passage, then some construct in addition to reading comprehension was being measured.

Katz and Lautenschlager (1995) identified factors that helped students answer the multiple-choice questions, such as the relative plausibility of the answer alternatives. They found that students selected the correct answer over the incorrect choices for more than 60% of the items they used. When insufficient information was available to answer a question (e.g., if the passage was missing), the incorrect choices should have been at least as likely as the correct one. This was obviously not the case. It would seem, these authors concluded, that the SAT reading task and others like it were psychometrically flawed instruments.

Freedle and Kostin (1994) argued that these multiple-choice reading tests do have construct validity. Past criticisms of multiple-choice tests have utilized regression analyses to demonstrate that item factors were more predictive of question difficulty than text factors (e.g.,

Drum, Calfee, & Cook, 1981). The plausibility of the alternatives, not the difficulty of the text, best predicted examinees' performance. Freedle and Kostin included a third category of predictor variables that reflected text-by-item overlap variables. Text-by-item overlap occurs, for example, when the question alternatives have words in common with the text. When these variables were included, only one of the eight significant predictors was an item variable; the other seven being text or text-by-item variables. The authors concluded that the students must have been utilizing text information in order to answer the questions, and that even though it is possible to answer many reading comprehension questions without the passage, as Katz, et al. have demonstrated, it is unlikely that students use this strategy when taking reading comprehension exams. More conclusive evidence concerning the construct validity of multiple-choice tests could be provided, Freedle and Kostin claim, if it could be demonstrated that examinees who perform well on these tests subsequently have a coherent representation of the text.

Multiple-choice tests have been attacked on other grounds. Available measures of reading comprehension (e.g., the SAT and the Armed Services Vocational Aptitude Battery) are very influenced by factors other than reading skill (Katz, Blackburn, & Lautenschlager, 1995; Katz & Lautenschlager, 1994; Katz, Lautenschlager, Blackburn, &

Harris, 1990; Pyrczak, 1972; van den Burg, 1990). Examples of these different factors are depicted in Figure 1.

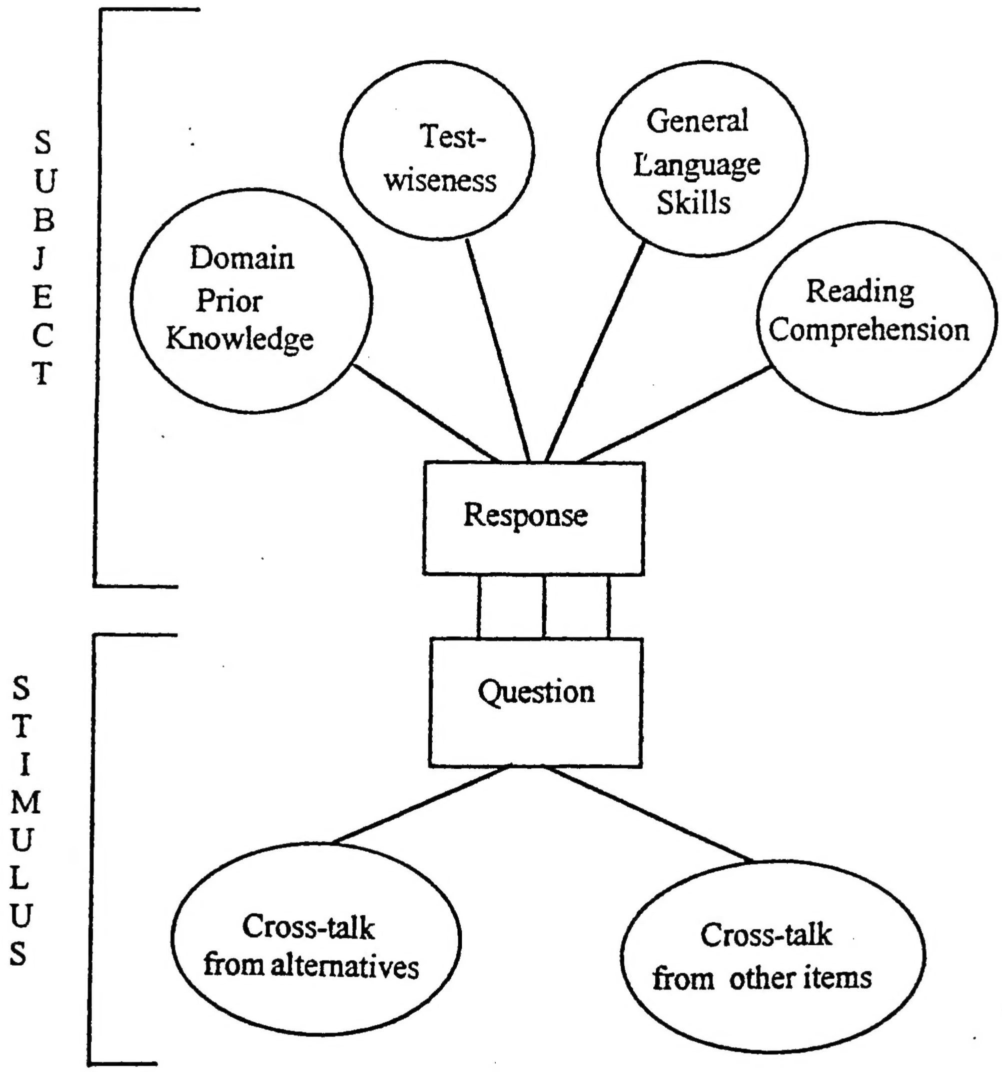


Figure 1. Examples of the Factors Influencing Reading Comprehension Measures.

These multiple-choice tests had three basic problems. The first is that students' prior knowledge was confounded with the measurement of comprehension (Johnston, 1984). This

made current reading comprehension tests, Johnston argued, merely good proxies for IQ. In fact, high levels of domain specific prior knowledge can obviate the need to read the text at all.

To deal with this problem, pretests of prior knowledge can be given. For example, content-relevant vocabulary tests can provide information about the examinee's prior knowledge, but means of objectively scoring these tests must be provided. Prior knowledge multiple-choice tests could also be given, but unless the domain specific concepts are selected very carefully, the potential for pretest sensitization is very high. Texts on obscure subject matters can also be used, but questions about external validity arise in such contrived tasks. The fact remains that there is no attempt made to assess readers' prior knowledge in current reading comprehension tests such as the SAT or the Armed Service Vocational Aptitude Battery (ASVAB).

It could be argued that reading comprehension can not, and should not be separated from the knowledge readers bring to a reading task, that is, prior knowledge. Acquiring knowledge is not so much acquiring new information, but connecting that new information to existing information (James, 1890). According to this view, reading comprehension involves integrating the information in a text into the existing knowledge structure of that text's domain. It is easier to understand a text for which we have much prior knowledge. Reading comprehension measures are used for

various purposes including to diagnose reading skill, assess a reader's knowledge, and to assess learning from text. The ability to measure prior knowledge without compromising post reading measures would aid in teasing apart these differences.

The second major problem of multiple-choice tests is that knowledge unrelated to either the text domain, or reading comprehension in general, is brought to bear in answering the multiple-choice questions on the reading tests. This general test taking ability, or test-wiseness, can greatly influence the readers' scores, as demonstrated by the effectiveness of coaching programs (e.g., Robinson & Katzman, 1986). Students have learned that many items of the test, as well as the alternative multiple-choices, give information that can be used to answer the questions. This cross-talk among items has been found to occur on 25% of the items on the SAT (Katz, et al.).

Problems one and two both involved confounding of reading comprehension with other knowledge and skills. A reader's score on a standard multiple-choice test, such as the SAT or Armed Services Vocational Aptitude Battery, consisted of that reader's accumulation of knowledge up to the time of testing, as well as a test-taking skill component. No attempt was made to compensate for the large individual differences that certainly existed in these components. Perhaps it was this fact that was responsible for the predictive validity of these types of tests. A score on a

multiple-choice test is reflective of the multiple-choice tests taken in the past, and predictive of multiple-choice tests to be taken in the future.

The third problem is the standard by which the student scores were measured. Different items on multiple-choice tests must use concepts that are distantly related to minimize cross-talk among the items. Furthermore, the items that discriminate good and poor readers were often items that were peripheral to the text (Johnston, 1984). As tests were refined, they contained more and more relatively trivial items (Tuinman, 1979). If the definition of reading comprehension included the formation of a coherent representation of the text, then this type of sampling strategy was very poor. The use of only divergent concepts and peripheral items could not assess the interrelationships among the concepts presented in the text; an assessment that was necessary if what was to be measured was the structure of the representation. The structures created during reading comprehension were ignored by reading comprehension tests.

Additionally, there is a single referent for a correct response on a multiple-choice test, and valid alternative answers could not be assessed. Presumably, there was an expert (or set of experts) that provided the answers for each question of a multiple choice test. However, experts do not always agree, and these differences are ignored by current multiple-choice tests. Utilizing these differences could have provided a more sensitive measure of reading skill in that

valid alternative perspectives on the relationships among the terms in a domain could have been used in diagnosing readers' understanding.

A better standard would sample as many of the most important concepts from the domain as possible, allowing assessments of the readers' comprehension structures to be made. Additionally, a referent for comparing the reader's structure that was sensitive to meaningful differences among the possible answers to the questions would provide more flexibility in assessment.

These problems with current multiple-choice tests of reading comprehension were related to the original concern; these tests were constructed atheoretically. What types of information did current theories of reading comprehension provide? Soon after we had learned to read, we begin to read to learn. In this sense, learning and comprehension are necessarily intertwined (Just & Carpenter, 1987). We read for entertainment and we read for knowledge. When we read to learn, we were told to read actively (e.g., Dansereau, Brooks, Holley, & Collins, 1978), that is, to consciously attend to and organize the information. This type of reading involved discovering connections and producing an organizational structure for use in later retrieval (Just & Carpenter, 1987). This involved constructing some mental model of the situation that included prior knowledge, the reader's goals, and the knowledge structure of the text. Multiple-choice reading comprehension tests asked the reader

to recognize concepts and facts without concern for the intricacy of the actual comprehension process.

The type of test used in this study, called the Cognitive Structure test, offers solutions for the above mentioned problems. The test is relatively novel and therefore helps to reduce the influence of differential strategies used by test-takers on the reading comprehension score, i.e, test-wiseness. The Cognitive Structure test samples concepts more densely from the domain and is also sensitive to meaningful differences between experts and thus sets a better standard by which comprehension can be measured. Additionally, the Cognitive Structure test can be used to investigate the structure of the examinee's knowledge and make diagnoses to repair misconceptions.

This study compared the Cognitive Structure test to two examples of standard reading skill tests; the Scholastic Aptitude Test (SAT) and the paragraph comprehension section of the Armed Services Vocational Aptitude Battery (ASVABpc). First, a brief history of cognitive structure testing was traced from early instantiations such as Shavelson (1972) to the current work directed by Bruce Britton. The work of Jonassen (1995) provided an excellent framework for directing this historical trek. Next, a simple example of the Cognitive Structure test is offered. Current theories of reading comprehension are reviewed and related to the types of information provided by both Cognitive Structure tests and current multiple-choice tests of reading skill. Empirical

investigations were used to address the relative construct and predictive validity of the two types of tests.

Cognitive Structure

Cognitive structures, and the novice-referent structure paradigm have been used in many instantiations by many researchers (e.g., Acton, Johnson, & Goldsmith, 1994; Britton & Gulgoz, 1991; Britton & Sorrells, in press; Britton & Tidwell, 1993; Medina-Diaz, 1993; Naveh-Benjamin, 1986; Shavelson, 1972, 1974, 1983).

Jonassen (1990) provided a good review of the history of cognitive structures, as well as a handbook for eliciting, representing, and assessing structural knowledge. His ideas are depicted in Figure 2.

The idea of knowledge having structure dates back (at least) to Aristotle. More modern epistemological theories have described knowledge as structured in schemas (e.g. Bartlett, 1932; Rumelhart, 1980), or semantic networks (e.g., Quillian, 1968; Collins & Quillian, 1969). Cognitive structures, which were also considered knowledge structures, were the representations in memory of the patterns of relationships among the concepts in a domain (Preece, 1976). Shavelson (1972) described cognitive structures as the organization of relationships among concepts. All of these theories, Jonassen pointed out, shared the important idea that knowledge was structured by the semantic relatedness of the concepts of the knowledge domain. This structure

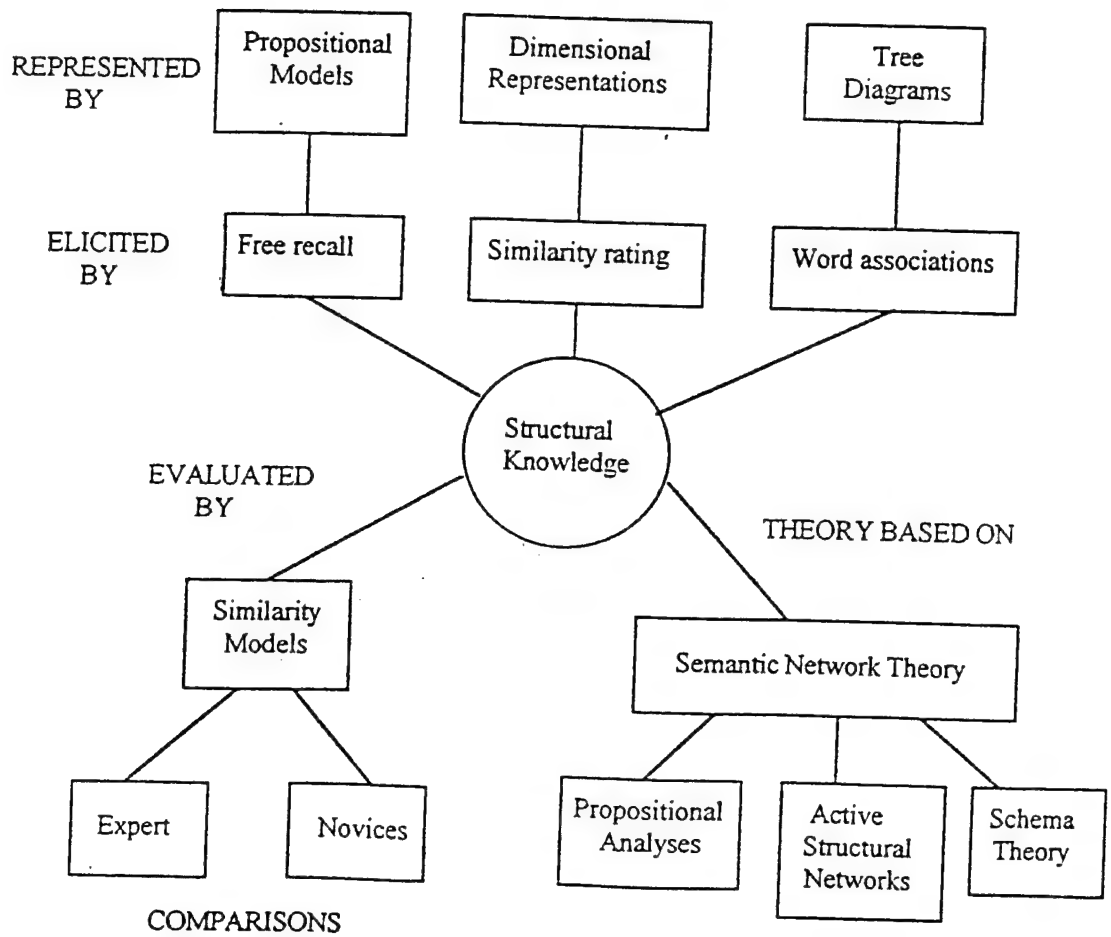


Figure 2. Jonassen's Ideas of Structural Knowledge

expedites the use of knowledge for solving problems in that cognitive structures, according to Jonassen, bridged knowing how (procedural knowledge) with knowing that (declarative knowledge). In this sense, the structures were not internal,

per se, but a product of the resources of the individual and the demands of the task.

Jonassen contended that the cognitive structures of experts were necessarily different than those of novices. Some examples of these differences were that experts were better than novices at perceiving meaningful patterns in their domain, at representing the problem on a deeper level, and at analyzing problems in a quantitative way. There was also a great deal of empirical evidence demonstrating these differences in several domains (e.g., Chase & Simon, 1973; Charnes, 1979; Chi, Glazer, & Farr, 1988;).

Cognitive structures can be elicited, or reified as Jonassen (1990) claimed, by a variety of techniques (see Figure 2). Relatedness ratings provided one way to spatially reify the construct of cognitive structure. Individuals were asked to rate pairs of terms (or concepts or ideas) on a fixed scale, similar to a Likert scale. These ratings were represented in a proximity matrix containing all the terms. Spatial representations of this matrix (e.g., Addtrees, multidimensional scaling solutions, or pathfinder networks) provided geometric distance information that was in turn interpreted as semantic distance information. These ratings were relatively reliable over time, and experts had a high degree of correspondence between their ratings (Diekhoff & Wigginton, 1989). Problems with relatedness ratings included defining relatedness to the examinee, and limiting the domain such that the number of concepts judged would remain

relatively few. For example, since the number of pairwise judgments equals n(n-1)/2, as few as 15 concepts required 105 comparisons. Fatigue and motivation must be considered in constructing the tests.

Another consideration in using relationship judgments as a means of assessing cognitive structure was the referent to which the structures were compared. Shavelson (1972) modeled the content structure of a physics text and compared students' cognitive structures to this referent both before and after instruction. He found that repeated exposure to the terms was not sufficient to change the cognitive structures of the students, but instruction resulted in cognitive structures that more closely resembled the content structure of the text. Other studies utilizing different domains have found that students' structures move closer to referent structures after instruction (e.g., Acton, Johnson, & Goldsmith, 1994; Gesslin & Shavelson, 1975). These studies, along with many others, provided good evidence that changes in cognitive structures can be used as reliable measures of learning (Naveh-Benjamin, McKeachie, Lin, & Tucker, 1986).

Diekhoff (1983) devised a way to use similarity judgments to compare students' cognitive structures to those of experts'. The ratings of 120 introductory psychology students were compared to a composite structure of the median responses of 10 subject matter experts. Pearson correlation coefficients were used to assess the correspondence between the structures. Multiple-choice and essay exams were also

given. Correlations among the tests revealed that the relationship judgments systematically varied with the other measures of knowledge. These results were also consistent with findings of other studies which employed a multidimensional scaling approach to representing the structures (e.g., Fenker, 1975; Johnson, Cox, & Curran, 1970; Wiener & Kaye, 1974). Diekhoff concluded that relationship judgments provide a valid assessment of knowledge. Furthermore, he pointed out that it is important to design testing methods that require students to consider the relations among concepts. This type of testing can promote, as well as be used to evaluate, structural knowledge.

Naveh-Benjamin, McKeachie, Lin, & Tucker (1986) used the ordered tree technique of Reitman and Rueter (1980) to assess the cognitive structures of 254 students, comparing them to the structures obtained from the course instructors. The student structures moved closer to that of their instructor during the course, and provided a valid means for differentiating different levels of achievement.

Evidence has been provided that these assessments of cognitive structures compared to some referent structure can be used to diagnose and repair misconceptions (e.g., Britton & Tidwell, 1995; Reif & Heller, 1982; Shavelson, 1972). Stanners and Brown (1982) gave students explicit information about the misconceptions revealed in their cognitive structures. This group of students subsequently exhibited

structures that were closer to the referent than to another group who were not given the explicit information.

Currently, work utilizing the relatedness-ratings paradigm is being conducted by and under the direction of Bruce Britton (Britton & Eisenhart, 1993; Britton & Gulgoz, 1991; Britton & Sorrells, in press; Britton & Tidwell, 1995; Tidwell, 1989; Tidwell, 1992). In these studies the Cognitive Structure Test has provided useful information about the knowledge structures of examinee's and experts, with this information used to test the effectiveness of rewriting instructional texts (Britton & Gulgoz, 1991), to diagnose and repair readers' misconceptions (Britton & tidwell, 1995; Tidwell, 1992), and to determine the optimal components for teaching and learning knowledge from particular domains (Britton & Sorrells, in press).

An Hypothetical Example of the Cognitive Structure Test

The Cognitive Structure test involves selecting the important terms or concepts from expository passages about various domains. Past research has demonstrated that experts agree on which terms from a passage are the important ones (Britton & Tidwell, 1995). Next, all possible pairs of the terms are presented to the examinee, with each pair accompanied by a six-point Likert scale. Figure 3 contains a simple text about the atom, with the most important terms arranged into three questions.

The examinee is instructed to give their best judgment as to the relatedness of the terms by selecting the number

Text:

The protons and the neutrons bind together, while the electrons orbit around them.

Terms: Proton, Neutron, Electron

Question #1

Proton Neutron

1	2	3	4	5	6
very related				very unrelated	

Question #2

Proton Electron

1	2	3	4	5	6
very related				very unrelated	

Question #3

Electron Neutron

1	2	3	4	5	6
very related				very unrelated	

Figure 3. A Hypothetical Example of the Cognitive Structure Test

that corresponds to their choice on the Likert scale. These responses are recorded in a matrix where each cell represents the relatedness of the terms from that column and that row. In the present study, the examinee's responses were compared (using Pearson correlation coefficients) to the responses of

each of 2 experts, or to the average of those two experts in some cases. The highest correlation between the examinee's responses and one of the two experts was converted to a Fisher z and recorded as that examinee's cognitive structure score for that passage.

Graphically representing the Cognitive Structures

Several techniques have been used to graphically represent the underlying structures derived from proximity data including Pathfinder analysis (Schvaneveldt, 1990), MDS (Kruskal & Wish, 1978; Wainer & Kaye, 1974; Weinberg & Menil, 1993), Addtree (Sattath & Tversky, 1977), and Extree (Corter & Tversky, 1986). Although there is some disagreement in the literature over which technique is best, all of these techniques have had success, and the various methods often yield similar results. All of these studies made the underlying assumption that knowledge has structure, and that this structure can be captured from matrices of associations (e.g., frequencies, co-occurrences, relatedness ratings, and correlations) collected from subjects.

One way to graphically represent the structures of examinees is illustrated in Figure 4. This Addtree graph represents the responses of an expert to the cognitive structure questions for a short passage, "Alkali" (see Appendix B for text). This passage is about the reactive alkali metals that form hydroxides and must be stored in inert substances like kerosene. The horizontal distances

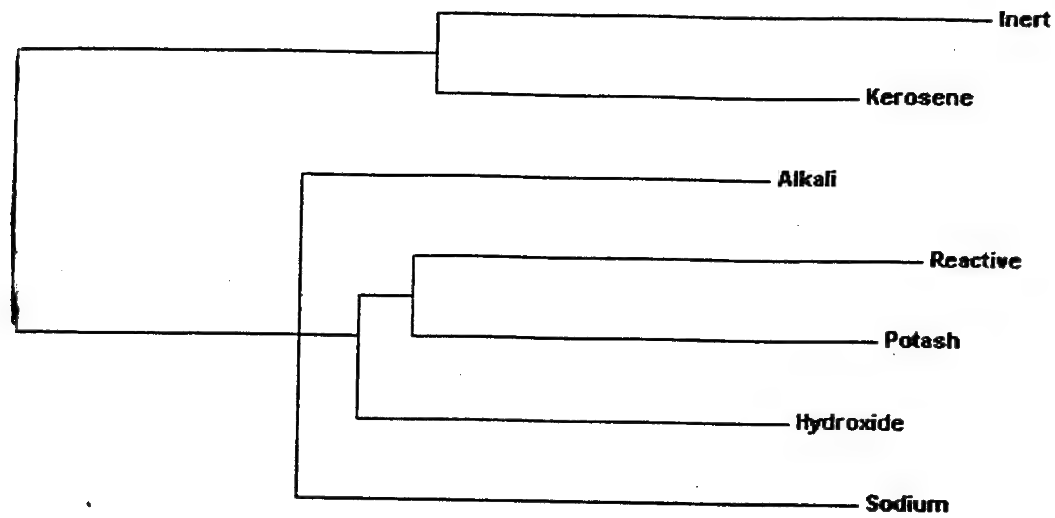


Figure 4. Addtree Graph for Alkali

between the terms in the graph represent semantic distance between the terms in the mental structure of the examinee.

Current theories of reading comprehension

Perhaps the most influential current model of reading comprehension is the Construction-Integration model of Kintsch (1988). This model emerged as a reaction to Kintsch's and van Dijk's earlier models (e.g., van Dijk & Kintsch, 1983) which relied on top-down, schema-driven processes to construct mental representations of the text that incorporated textbase representations with prior knowledge to produce a situational model. The textbase model consisted of the propositions derived from the text, along with certain so-called backward, or anaphoric inferences which maintain referential coherence. The Construction-Integration model adopted a weaker account of the role of prior knowledge in

order to construct a more flexible and adaptive model (Kintsch, 1988, 1992).

The Construction-Integration model utilized a generic connectionist framework (e.g., McClelland & Rumelhart, 1986) where input propositions were used to create the nodes of the semantic network. Links were created between the nodes that were embedded (i.e., propositional embedding) or that shared arguments. Following the generic connectionist framework, the nodes were given some initial activation, the activation spread and was modified through a connectionist algorithm, and iteration of this process continued until the network settled. Settling occurred when the activation vectors ceased to change at some preselected criterion. This process was said to occur in a fixed-capacity working memory, with the most active propositions (usually only two) held over for integration in the next processing cycle (usually the next sentence). When propositions with activations too low to be held over were needed to maintain coherence, a reinstatement search in long-term memory was carried out. Reading comprehension, according to this model, was the product of two processes: constructing the propositional network (construction), and subsequently editing that network (integration). This network has structure, and was constructed and modified in working memory.

Just and Carpenter (1987, 1992) also viewed reading comprehension as consisting of several levels of representation along with each level's associated processes.

These processes operated in parallel, were mostly automatic, and required management of working memory. Reading comprehension, according to Just and Carpenter, could be best understood in a larger information processing context called the Collaborative, Activation-based Production System (CAPS). The CAPS architecture was similar to symbolic, or information processing approaches in that processing occurs in discrete stages according to rules called productions. These productions were contingency statements used for example, to detect a noun phrase. During comprehension, encountering the word "the" would signal a noun phrase - the reader would assume that a noun phrase would follow. However, additional productions simultaneously competed for activation. When the conditions for a particular production were satisfied, the actions were performed (i.e., the production fires). This conflict resolution between productions was the collaborative aspect of the model. Activations could exist at several levels. When the activation of a production reached its threshold it fired. The knowledge structures created in working memory during comprehension were modified by the productions. Subsequently, a new cycle began in which the knowledge structure may satisfy different productions. The system moved from state to state, but concurrent productions can fire in parallel. This model was often referred to as a hybrid connectionist-symbolic model because of these two types of processing. As in Kintsch (1988), the ideas of

knowledge having structure and dynamic management of working memory were central to the theory of reading comprehension.

Just and Carpenter modified their CAPs approach to accommodate working memory limitations (1992). Concurrent demands on working memory imposed a limitation on the amount of activation allocated to the production system. In this sense, working memory does not have a fixed buffer, per se, but was limited by processing demands on the entire working memory system. This new model was termed Capacity Constrained Collaborative Activation-based Production System (3CAPS). Reading comprehension remained a process of constructing coherent representations in working memory, but was also influenced by concurrent demands on the processing system as a whole.

An attempt was made to embed the Construction-Integration model into capacity constrained CAPs architectures (Goldman, Varma, & Cote, 1996). This allowed for several propositions to be held over in working memory when demands on this resource were low. The result was a more richly connected network that preserved more of the textbase information. Multiple representations can be activated (and compete) allowing other processes necessary for constructing the situation model (i.e., prior knowledge activation) to facilitate comprehension. As with the other theories, structured knowledge, dynamic management of working memory, and integration of prior knowledge with textbased structures were central to this view of reading comprehension.

Gernsbacher (1990, 1996) has proposed that reading comprehension has as its goal the building of coherent mental structures. Three processes and two subsequent mechanisms drive this structure building. First, foundations are laid upon which to build the structure. This process requires much cognitive effort, evidenced by the relatively slow processing of initial discourse information. The second and third processes involve mapping information onto the developing structures, and shifting initiative to building new substructures. These structures are modified by the mechanisms of enhancement and suppression, mechanisms which are very similar to the backpropegation algorithms of the connectionist approaches.

Graesser and Britton (1996) identified several metaphors that are pervasive in current models of text comprehension. These metaphors can be combined into the following working definition of text comprehension: Comprehension is an active management of working memory where multi-level coherent representations and inferences are constructed (p 350). The idea of comprehension as structure building, although not new, is ubiquitous in theories of reading comprehension (e.g., Gernsbacher, 1990; Kintsch, 1988; van Dijk & Kintsch, 1983). If current reading comprehension tests were constructed atheoretically as Just and Carpenter claimed, then a test that is based on theory and that assesses the comprehender's structure is needed.

Statement of the problem

It has been demonstrated that standard multiple-choice tests of reading comprehension have poor construct validity. They ignore valuable insights into the process of comprehension that have been provided by current theories of reading comprehension. Understanding a text involves creating a coherent representation, but standard tests make no attempt to assess this structure. This study compared the construct and predictive validity of multiple-choice testing from a version of the SAT and the paragraph comprehension section of the Armed Services Vocational Aptitude Battery (ASVABpc) to Cognitive Structure testing of the same passages. It was predicted that the Cognitive Structure test would demonstrate better construct validity than the multiple-choice tests.

Construct Validity. Nunnally (1978) contended that construct validity, unlike the measures of reliability and predictive validity, can not be directly measured. He then prescribed three major aspects of construct validity that could be used to assess new measures. The first was identifying the domain variables (i.e, observables) related to the construct of interest. Second, the extent to which the variables measured the same thing must be determined. Third, the extent to which the measures of the construct produced results (i.e, systematic variance) which were predictable from hypotheses about accepted theoretical constructs must be determined. The first aspect was a theoretical concern and draws upon the theories reviewed earlier as well as upon

informed intuitions about the nature of reading comprehension, whereas aspects two and three were statistical and were addressed in the Method section.

The theoretical aspect (covered in the Introduction) involved defining the construct, as well as specifying the expected relationships among the measures of that construct. Reading comprehension was described as the process of constructing in working memory coherent representations of the text knolwedge, and integrating those representations into semantic memory via text and situational based (i.e., information from the text combined with prior knowledge) inferences. Measures of reading comprehension should consider these processes. Additionally, measures of reading comprehension should be inter-correlated; they should vary systematically with similar constructs and measures of reading skill (convergent validity), and vary unsystematically with dissimilar constructs such as measures of mathematical reasoning (divergent validity). These theoretical considerations provided information about what to do in the second and third aspects, which statistically assessed to what extent the constructs in question (the reading comprehension tests) measured the same thing.

To investigate the second aspect, a correlation matrix comprised of the Cognitive Structure Test and the multiple-choice tests, along with several other types of general and specific knowledge tests provided insight as to what the different tests measured. Very high correlations between the

Cognitive Structure Test and the multiple-choice tests would indicate that the tests measured the same construct in the same way. If this proved to be the case, a statistical argument could not have been made for using the Cognitive Structure Test, since more established tests can be used to measure reading comprehension. However, other arguments (e.g., theoretical or pragmatic) may also affect the choice of test. Moderate correlations between the multiple-choice tests and the Cognitive Structure test would suggest that the tests measured different (albeit related) constructs. Low correlations would be interpreted as low or no overlap in the constructs to be measured. It was predicted that the correlations would be moderate, suggesting that the constructs measured were similar, but not identical.

The patterns of correlations of the Cognitive Structure and the multiple-choice tests with other tests of general and specific knowledge were investigated to assess the convergent and divergent validity aspects of the tests' construct validity. It was predicted that these patterns would be very similar in that both tests would be moderately to highly correlated with the tests that have strong verbal components, and not correlated with the math and vocational abilities tests.

Factor analyses of these tests provided similar information about the convergent and divergent validities of the Cognitive Structure and multiple-choice tests. It was predicted that the Cognitive Structure and multiple-choice

tests would be more similar to measures of reading skill than to measures of vocational skill or mathematical ability.

Nunnally's third aspect was carried out by splitting the data files at median levels for another measure of reading comprehension (e.g., general knowledge verbal) and comparing how the Cognitive Structure Test and the multiple-choice tests varied with this manipulation. To demonstrate construct validity, the recruits who scored high on the general knowledge verbal measure should also have high scores on the Cognitive Structure and multiple-choice measures (and the same trend for the low scores). It was predicted that the variance in scores on the Cognitive Structure Test would be more similar to the variance in the general knowledge verbal measure of reading skill, than the multiple-choice tests.

<u>Predictive Validity.</u> Regression procedures were used to evaluate the predictive validity of each test for various criteria. It was hypothesized that the Cognitive Structure Test would have greater predictive ability than the multiple-choice tests. The strategy of splitting the files used in Aspect 3 of the construct validity analyses was also used here to assess predictive validity as a function of the recruits' general knowledge verbal scores. It was predicted that, for both the high and low ability readers (as assessed by the general knowledge verbal scores), the Cognitive Structure Test would be a better predictor than the multiple-choice tests for the various criteria.

These statistical comparisons were combined with theoretic and pragmatic concerns (e.g., cost of construction and updating) to provide arguments relevant to adopting the Cognitive Structure test as an improved measure of reading skill.

Two other issues were also addressed concerning the use of the Cognitive Structure test. The first issue involved comparing the Cognitive Structure and multiple-choice test responses of the experts. The second involved graphically representing the knowledge structures derived from the Cognitive Structure Test in two domain.

Method

Subjects

Recruits. 347 Air-force recruits participated. Recruits with more than 3 missing values for each Cognitive Structure test (one test for each of the 16 passages) were eliminated.

Experts. 32 subject matter experts (see Appendix A) voluntarily participated. 28 were University faculty members with a Ph.D. in the domain from which the texts were taken (e.g., physics, biology, chemistry). The 4 expert participants who did not hold a Ph.D. at the time of testing were career individuals in the fields corresponding to the domains of passages they read. All experts were recruited personally, and agreed to participate.

Materials

Texts: 16 passages, 8 from the Scholastic Aptitude Test (SAT) and 8 from the paragraph comprehension section of the Armed Services Vocational Aptitude Battery (ASVABpc), were used. The SAT passages were longer than the ASVABpc, and were accompanied with 3-5 multiple-choice questions each, while the shorter ASVABpc passages had 1 multiple-choice question each. (The terms used in the Cognitive Structure Test of each passage are shown in Appendix B along with the multiple-choice questions, and the passages themselves.) Every subject read all passages and took all tests.

Cognitive Structure Tests: Two different versions of each Cognitive Structure Test were constructed. One version of the Cognitive Structure Test used the important terms from the passages and asked participants to rate all pairs of these terms on a six point Likert scale. This version was always administered after reading the passage, and before reading for one-half of the passages as a prior knowledge measure. This version was called the Cognitive Structure Test. The other version used different terms taken from the same domain as the text, but that did not appear in the text itself. This version was administered only before the passages were read, and will be called a Prior Knowledge Cognitive Structure Test to distinguish it from the cognitive Structure Test described above. (These tests can also be found in Appendix B.)

Cognitive Structure scores: For each passage, the Cognitive Structure Test was taken by two experts. The recruits' responses (relatedness ratings) were correlated with each expert's responses such that each recruit had a Pearson r correlation coefficient for expert 1 and expert 2. In order to conduct inferential statistical tests about these correlations, these coefficients were converted to Fishers z's. Each participant's Fishers z for expert 1 and expert 2 were compared and the highest z was taken as that participant's Cognitive Structure score.

Other tests: Batteries of general and specific domain knowledge were also administered to the recruits. These tests included the Nelson-Narens general knowledge subtests for verbal knowledge (general knowledge verbal: GKV) and a spatial ability test provided by Pat Kylloyen that assessed general awareness of direction and movement (general knowledge spatial: GKS) as well as various subtests of the Armed Services Vocational Aptitude Battery (ASVAB) intended to measure knowledge in various specific domains.

The ASVAB is a group of tests for measuring general ability as well as specific abilities believed to be relevant in making decisions about selection and placement in the military. The subtests used in the present study were: Math Knowledge, a subtest measuring knowledge of high school mathematics principles; Paragraph comprehension, a subtest measuring the ability to obtain information from written passages; Word Knowledge, a subtest measuring the ability to

select the correct meaning of words, and to select appropriate synonyms; General Science, a subtest measuring knowledge of the physical and biological sciences; Arithmetic Reasoning, a subtest measuring the ability to solve arithmetic word problems; Electronics Information, a subtest measuring knowledge of electricity and electronics; Autoshop, a subtest measuring knowledge of automobiles, tools, and shop terminology and practices; and Mechanical Comprehension, a subtest measuring knowledge of mechanical and physical principles and the ability to visualize how illustrated objects work. A composite test called the Armed Forces Qualification test (AFQT) from the ASVAB was used as a general ability measure [AFQT = 2(Paragraph Comprehension + Word Knowledge) + Math Knowledge + Arithmetic Reasoning].

Each recruit was tracked through completion of their Armed Services Technical Schooling. There are two divisions of this schooling; general and special training. The final grade averages of the recruits participating in the general part of the training school were provided for use as an external criterion of ability. This general training was intended to provide the recruit with the basic knowledge and skill necessary to perform in the Armed Services. Examples of the courses in this school included Airborne Communications, Operations Resources, Aircrew Life Support, and Weather.

<u>Design and Procedure</u>

The recruits were presented all passages via a computer monitor. A within-subject design was used, and all passages

and tests were randomized in a counterbalanced order. They were given instructions as to the nature of the task as well as how to respond. In general, the Cognitive Structure tests presented all possible pairs of the important terms associated with each passage, and asked the participant to provide a response on a 1 to 6 point scale, indicating their best judgment about the correct relation between the members of each pair of terms.

The recruits were allowed to look back at the passages while taking the test (Comprehension condition) for one-half of the passages, and not allowed to look back (Memory condition) for the other half. The counterbalanced prior knowledge tests consisted of the same terms as the Cognitive Structure tests for one-half of the presented passages, and of different terms from the same domain for the other half of the passages. The multiple-choice questions were always administered after the passages, with half of the passages accompanied by the multiple-choice test first, then the Cognitive Structure test, and half of the passages with the Cognitive Structure test coming first, then the multiple-choice test. This resulted in eight counterbalancing conditions (see Table 1 for examples of these conditions); 2 (comprehension vs. recall) X 2 (order of prior knowledge tests) X 2 (order of post tests).

The experts were given the Prior Knowledge Cognitive Structure Test (different terms than those in the passage) first, then asked to read the passage, and then completed the

Table 1

Counterbalancing Conditions for the ASVABpc Passages

Comprehension Condition

Condition	Prior Knowledge	Passage*	Post Tests	
1	PK	Alkali	MC	CS
2	PK	Automobile	CS	MC
3	CS	Inflation	MC	CS
4	CS	Lightning	CS	MC

Memory condition

Condition	Prior Knowledge	Passage	Post Tests	
5	PK	Car and curve	MC	CS
6	PK	Smoke	CS	MC
7	CS	Tactile	MC	CS
8	CS	Wood	CS	MC

Note: PK = Prior Knowledge Cognitive Structure Test with terms not in passage. CS = Cognitive Structure Test with terms from passage. *The order of these passages was random for each recruit.

Cognitive Structure test and the multiple-choice tests in a random order.

Analyses

General Descriptives and Comparisons. Descriptive statistics were generated for all variables. One-way ANOVAs and post-hoc comparisons were conducted to investigate potential differences between the counterbalanced orders (eight orders) of the pretests and posttests, between the Cognitive Structure scores for the ASVABpc vs. the SAT passages, and for differences between the multiple-choice scores for the ASVABpc and the SAT passages. Planned contrasts were also conducted to investigate potential differences in the Cognitive Structure scores and the multiple-choice scores for the Memory and the Comprehension conditions. Further analyses were used to compare the relative construct and predictive validity of the Cognitive Structure tests and the standard multiple-choice tests.

Construct Validity. The guidelines prescribed by Nunnally (the statistical aspects described earlier) for assessing construct validity were used. Correlations were computed between the Cognitive Structure tests and the multiple choice tests, as well as among the other subtests. Two separate factor analyses were computed for these tests; one including the Cognitive Structure and the multiple-choice scores, and one excluding these variables. The data files were split at median performance levels for the general

knowledge verbal measure, and the variance in the Cognitive Structure and multiple-choice scores were compared.

Predictive Validity. Regression models were set up to determine the predictiveness of the Cognitive Structure test vs. the multiple-choice tests on three classes of criteria; external criteria (the Technical School scores), general knowledge criteria (General Knowledge Verbal and General Knowledge composites), and domain specific criteria (General Science, Autoshop, Electrical Information, and Mechanical Comprehension). The reported standardized beta coefficients were interpreted as rough indices of the influence of each variable on the model.

Additionally, the data were divided into subsets of passages that corresponded to the paragraph comprehension section of the ASVAB, and to passages that corresponded to the SAT. The above analyses were conducted on the entire data set, the ASVABpc subset, the SAT subset, and for the individual passages.

The Experts. Differences among the experts were investigated. Presumably, there is an implicit expert deriving the correct answers for the multiple-choice tests. In other words, someone (presumably a subject expert) has determined which of the alternative choices correctly answers the question. Therefore, comparisons were made between the answers provided by the implicit experts, and the answers provided by the experts in this study. Correlations among the Cognitive Structure Test responses of the different subject

experts were computed, and comparisons of performance on the multiple-choice tests were made.

Graphical Representations of the Structures. Two passage were selected to demonstrate two different types of graphical representations, Addtree (Sattath & Tversky, 1977) and Extree (Cotter & Tversky, 1986), and the potential diagnostic aspects of the Cognitive Structure Test.

All analyses were conducted on an IBM clone PC using the Statistical Package for the Social Sciences.

Results and Discussion

General Descriptives. Descriptive statistics for the Cognitive Structure and multiple-choice tests were presented in Table 2. In general, the SAT multiple-choice scores (M=29.82, sd=13.25) were lower than the ASVABpc multiple-choice scores (M=59.57, sd=21.29), $t_{(289)}=27.11$, $p<.001$. One of the eight SAT passages had a multiple-choice average that were not significantly greater than chance (Jetstream) while all of the ASVABpc multiple-choice averages were greater than would be expected by chance. The SAT Cognitive Structure scores (M=.32, sd=.20) were also lower than the ASVABpc Cognitive Structure scores (M=.50, sd=.20), $t_{(248)}=13.67$, $p<.001$, and were all significantly greater than chance.

Cronbach's alphas indicated that the two tests were very similar in reliability. The Cognitive Structure Tests' alpha was .63 and the multiple-choice tests' alpha was .68. This was lower than other reports of alphas for multiple-

Table 2

Means and SDs for CS Scores and MC Scores by Passage

Text	#Q	MCMean(%)	SD	CSMean	SD
Alkali	1	68.48	46.52	.45	.44
Automobiles	1	79.07	40.73	.50	.37
CarCurve	1	52.33	50.00	.41	.43
Inflation	1	49.35	50.06	.62	.71
Lightning	1	90.16	29.83	.98	.63
Smoke	1	40.42	49.14	.24	.38
Tactile	1	50.39	50.06	.27	.35
Wood	1	43.80	49.68	.40	.37
ASVAB Combined	8	59.30	21.28	.50	.19
Jetstream	5	20.16*	19.32	.14	.28
Bacteria	4	26.49*	25.12	.45	.61
Theories	5	26.75*	20.68	.29	.42
Prostag	4	29.12	25.64	.49	.43
Climate	3	30.84	26.98	.30	.35
Glaciers	5	32.56	22.22	.31	.34
Radios	5	34.83	24.38	.10	.25
Surfaces	4	35.01	25.22	.35	.31
SAT Combined	35	29.70	13.12	.31	.21

Note. #Q = Number of questions accompanying passage. *Not significantly different than chance.

choice tests (like the SAT) perhaps because only 8 ASVAB and 8 SAT passages were used, i.e., they were not complete tests.

There were no effects for condition (memory vs. comprehension) in the combined data set for either the Cognitive Structure scores or the multiple-choice scores. However, the subset of ASVABpc passages yielded lower multiple-choice scores for the memory condition, that is, when the participants could not refer back to the passage (M=55.28, sd=20.66) than for the comprehension condition where participants could look back (M=62.37, sd=21.14), $t_{(326)}$= -3.06, $p<.003$.

There were no effects for test order. Neither the Cognitive Structure nor multiple-choice scores were systematically affected by the Prior Knowledge Cognitive Structure Test or the Cognitive Structure test administered as pretests. This indicated that the prior knowledge measures can be taken with the Cognitive Structure tests without biasing post reading measures.

The SAT passages were more difficult than the ASVABpc passages, regardless of the test used to measure their difficulty. Both the Cognitive Structure scores and the multiple-choice scores were lower for the SAT passages. The chance score for the one SAT passage, and the overall low performance on these passages could account for the fact that there was no difference between the memory and the comprehension conditions, i.e., the participants in this condition were guessing and did not benefit from being able

to look back at the passage. When the participants read the ASVABpc passages, the ability to look back at the text resulted in higher multiple-choice scores.

The descriptive statistics for the ASVAB subtests and the Nelson Narens General Knowledge Verbal test are in Table 3.

Construct Validity

Correlations. The correlations between the Cognitive Structure and multiple-choice tests are in Table 4. When all passages were combined, the correlation between the Cognitive Structure test and the multiple-choice test was $\underline{r}_{(249)}=.666$, $\underline{p} <.001$. The correlation in the ASVABpc passages between the Cognitive Structure and multiple-choice scores was $\underline{r}_{(287)}=.387$, $\underline{p}<.001$, and in the SAT passages the $\underline{r}_{(295)} = .682$, $\underline{p}<.001$.

The correlations of the Cognitive Structure and multiple-choice scores with the other variables presented in Table 5 were arranged in the combined column such that the highest correlations with the Cognitive Structure Test were at the top descending to the lowest correlations at the bottom. In this column, the patterns of correlations for the Cognitive Structure and multiple-choice tests are nearly identical.

For the SAT passages, the patterns of correlations were also very similar with both the Cognitive Structure and multiple-choice tests most highly correlated with the AFQT, General Science, General Knowledge Verbal, and Word Knowledge.

Table 3

Means and Standard Deviations for All Tests

Test	Mean	SD
Autoshop	50.35	8.25
ElectricInformation	51.35	7.39
GeneralScience	53.37	6.84
ArithReasoning	54.10	6.81
MechanicalComp	53.98	8.16
WordKnowledge	54.43	4.58
ParagraphComp	55.01	4.11
MathKnowledge	56.13	7.00
GKVerbal	62.44	15.71
TechnicalSchool	86.26	5.68
AFQT	219.88	16.49

Note. GKVerbal = General Knowledge Verbal. AFQT = Air Force Qualifications Test (see text for details).

Table 4

Correlations of Cognitive Structure Scores with the Multiple-Choice scores by Passage

ASVABpc	Correlation	SAT	Correlation
Alkali	.07	Bacteria	.34*
Automobiles	.10	Climate	.34*
Inflation	.19*	Glaciers	.36*
Lightning	-.02	Jetstream	.08
Smoke	.06	Theories	.22*
Tactile	.14*	Prostagland	.28*
Car in the curve	.01	Radios	.09
Wood	.26*	Surfaces	.34*

Combining Passages	Correlation	R^2
ASVABpc	.39*	.15
SAT	.68*	.46
All Passages	.67*	.45

Note. * Correlation is significantly different than zero (df=346).

Table 5

Correlations of the Cognitive Structure Test and the Multiple-Choice Tests with the Controls by Passage Type

	ASVABpc		SAT		Combined	
	CS	MC	CS	MC	CS	MC
AFQT	$.35^{1}$	$.39^{1}$	$.47^{2}$	$.56^{1}$	$.51^{1}$	$.57^{1}$
GKV	$.32^{3}$	$.35^{2}$	$.48^{1}$	$.50^{3}$	$.49^{2}$	$.52^{2}$
WORDKNOW	$.29^{6}$	$.35^{3}$	$.47^{3}$	$.51^{2}$	$.46^{3}$	$.52^{3}$
GENSCI	$.32^{4}$	$.32^{5}$	$.45^{4}$	$.49^{4}$	$.46^{4}$	$.51^{4}$
TECHSCHO	$.33^{2}$	$.24^{9}$	$.34^{6}$	$.37^{7}$	$.41^{5}$	$.39^{8}$
ARITHRES	$.29^{7}$	$.31^{7}$	$.33^{7}$	$.41^{5}$	$.39^{6}$	$.43^{5}$
ELECINFO	$.30^{5}$	$.31^{8}$	$.35^{5}$	$.37^{8}$	$.38^{7}$	$.40^{7}$
MECHCOMP	$.26^{8}$	$.33^{4}$	$.33^{8}$	$.38^{6}$	$.34^{8}$	$.41^{6}$
MATHKNOW	$.21^{10}$	$.22^{10}$	$.23^{10}$	$.29^{10}$	$.29^{9}$	$.31^{11}$
PARACOMP	$.18^{11}$	$.18^{11}$	$.28^{9}$	$.33^{9}$	$.28^{10}$	$.32^{9}$
AUTOSHOP	$.22^{9}$	$.32^{6}$	$.23^{11}$	$.27^{11}$	$.25^{11}$	$.32^{10}$

Note. The superscripts indicate the rank order of the correlations. All p's < .01. Degrees of Freedom for CS = 287, MC = 330.

For the ASVABpc passages, the patterns were slightly different. The top correlations with the Cognitive Structure Test were (in descending order) the composite AFQT, the Technical School scores, General Science, General Knowledge Verbal, Electrical Information, and Word Knowledge, and Arithmetic Reasoning. The multiple-choice scores (df = 330, $p<.001$) were most highly correlated with the AFQT, General Knowledge Verbal, Word Knowledge, Mechanical Comprehension, General Science and AutoShop.

Factor Analyses. Correlations among the subtests and the subsequent factor analyses solutions are presented in Table 6 and 7 respectively. Three factors emerged (Principal Components Analyses with Varimax rotation of factors). When the Cognitive Structure scores and the multiple-choice scores were excluded from the analyses, the first factor was comprised of Autoshop, Electrical Info, and Mechanical Comprehension. The second factor included General Science, Paragraph Comprehension, Word Knowledge, and General Knowledge Verbal. The third factor included Arithmetic Reasoning, Math Knowledge, and the Technical School scores. Including the Cognitive Structure scores and the multiple-choice scores in the analyses resulted in the same groupings of the tests, except that the Factor 2 tests moved to Factor 1 and included the Cognitive Structure scores, the multiple-choice scores, and now the Technical School scores.

The inclusion of the Cognitive Structure and multiple-choice scores seems to have revealed a general verbal factor

Table 6

Correlations Among Subtests (Controls)

	AF	AR	AS	EI	GS	MK	ME	PC	GV	TS	WK
AF	1.00	.81	.21	.34	.56	.76	.46	.46	.56	.40	.67
AR		1.00	.22	.24	.36	.56	.41	.25	.35	.33	.30
AS			1.00	.62	.34	.04*	.55	.07*	.44	.21	.22
EI				1.00	.46	.13	.59	.19	.55	.20	.39
GS					1.00	.33	.41	.30	.63	.36	.55
MK						1.00	.36	.15	.30	.31	.20
ME							1.00	.13	.49	.27	.30
PC								1.00	.22	.26	.34
GV									1.00	.32	.59
TS										1.00	.21
WK											1.00

Note. AF=AFQT. AR=Arithmetic Reasoning. AS=Autoshop. EI=Electrical Information. GS=General Science. MK=Math Knowledge. ME=Mechanical Comprehension. PC=Paragraph Comprehension. GV=General Knowledge Verbal. TS=Technical School Final Grade. WK=Word Knowledge. * $p > .05$ (NS).

Table 7

Factor Analyses of Tests

Rotated Matrix Without CS and MC Scores

Factor	Eigenvalue	Percent of Variance Explained
1	3.98	39.8

Test	Loading Factor
AUTOSHOP	.85
ELECTRICINFO	.81
MECHANICALCOMP	.74

Factor	Eigenvalue	Percent of Variance Explained
2	1.42	14.2

Test	Loading Factor
GENERALSCIENCE	.61
PARAGRAPHCOMP	.70
WORDKNOWLEDGE	.78
GKVERBAL	.57

Factor	Eigenvalue	Percent of Variance Explained
3	1.07	10.7

Test	Loading Factor
ARITHREASONING	.79
MATHKNOWLEDGE	.88
TECHNICALSCHOOL	.50

Table 7

Factor Analyses of Tests

Rotated Matrix Including the CS and MC Scores

Factor	Eigenvalue	Percent of Variance Explained
1	5.01	41.7

Test	Loading Factor
GENERALSCIENCE	.59
PARAGRAPHCOMP	.68
TECHNICALSCHOOL	.42
WORDKNOWLEDGE	.75
GKVERBAL	.58
COGNITIVE STRUCTURE	.66
MULTIPLE-CHOICE	.68

Factor	Eigenvalue	Percent of Variance Explained
2	1.42	11.9

Test	Loading Factor
AUTOSHOP	.86
ELECTICINFO	.80
MECHANICALCOMP	.73

Factor	Eigenvalue	Percent of Variance Explained
3	1.09	9.1

Test	Loading Factor
AITHREASONING	.76
MATHKNOWLEDGE	.89

indicating that the variance in these tests was more similar to the variance in the other measures of reading skill (general knowledge verbal, paragraph comprehension, and word knowledge) than to the measures of math and vocational knowledge.

Interpreting the Correlations For Construct Validity.

Two separate issues are involved in interpreting the implications of the correlations for construct validity. First, does each test demonstrate construct validity when considered individually? What particular constructs are measured by the test, and are the constructs the same when the test is used in different contexts? This issue can be addressed by observing the ranked patterns of correlations (from Table 5) for each test in each condition (ASVABpc vs. SAT). The second issue involves the similarity in the constructs measured and the similarity in the way these constructs are measured by the tests. Do the Cognitive Structure Test and the multiple-choice test correlate with these constructs in the same way? This issue can be addressed by observing the amount of overlap in variance, as measured by the R^2's, between the two types of tests. The magnitude of the correlation between the Cognitive Structure Test and the multiple-choice test indicates how similarly the constructs are measured by the two tests.

Considering first the construct validity of the multiple-choice test, the ranking of correlations in the SAT passages and the ASVAB passages are similar, indicating that

this test measures the same constructs (e.g., AFQT, GKV, and Word Knowledge) in both passage types.

Considering the construct validity of the Cognitive Structure Test, the ranking of correlations are different for the different passages, indicating that this tests measures different constructs in the different passage categories. Most notably, the Cognitive Structure Test is correlated .47 (ranked 3) with WordKnowledge in the SAT condition, but only .29 (ranked 6) in the ASVABpc condition. Also, the Technical School scores are ranked 6 for the SAT passages and 2 for the ASVABpc passages. It seems that the Cognitive Structure Test measures verbal factors more in the SAT condition than in the ASVABpc condition.

In the combined passages, the rankings of correlations for the Cognitive Structure Test and the multiple-choice test are nearly identical, indicating that when the passages are combined, the tests measure very similar constructs. The fact that the tests measure similar constructs in the combined condition, but different constructs when the passage type is considered is reminiscent of the finding from the R^2's in Table 4; the Cognitive Structure Test and multiple-choice test have more overlap in variance in the combined and SAT conditions than in the ASVABpc condition. What are the implications of this for the construct validity of the two tests?

The problem is that each test has demonstrated construct validity on its own, but yet the two tests seem to

measure these constructs differently in the two types of passages. One factor that might be a precondition for establishing construct validity is that the passages are appropriate for the population being tested. The ASVABpc passages were designed to measure reading comprehension in military recruits, which is the population from which these participants were drawn. Therefore, the level of difficulty of the ASVABpc passages and their accompanying multiple-choice questions were likely to have been within the range of the participants' ability. In contrast, the SAT passages and questions were empirically more difficult than the ASVABpc passages and questions. The difference between the SAT and ASVABpc multiple-choice scores was significant ($t_{(289)}=27.11$, $p<.001$), as was the difference in Cognitive Structure scores ($t_{(248)}=13.67$, $p<.001$). The explanation of the fact that the two tests individually demonstrate good construct validity, but were not very correlated in the ASVABpc condition could perhaps reside in these differences in the difficulty levels of the passages.

Verbal ability differences among the recruits could also be responsible for this difference. To investigate this idea, the recruits were divided into high and low ability groups by splitting the file at the median level for general knowledge verbal (GKV). This measure was chosen because it loaded high on the verbal factor (.58) and was an external test (i.e., not part of the ASVAB or SAT). Table 8 contains the means, standard deviations and R^2's for the upper and

lower groups of recruits defined by the median split. For the higher ability readers (GKV > 62.5), the ASVABpc multiple-choice score mean was 66.58, and the SAT multiple-choice score mean was 35.01 ($t_{(137)}$=18.99, p<.000). For the lower ability readers (GKV < 62.5), the ASVABpc multiple-choice score mean was 53.21, and the SAT multiple-choice score mean was 25.11 ($t_{(151)}$=19.47, p<.001). As in the combined data, both the higher and lower ability readers had much difficulty with the SAT passages.

The Cognitive Structure scores reflected this same general trend. For the higher ability readers, the ASVABpc Cognitive Structure score (.56) was significantly higher than the SAT Cognitive Structure score (.39), $t_{(117)}$=8.38, p<.001. For the lower ability readers, the ASVABpc Cognitive Structure score (.44) was also higher than the SAT Cognitive Structure score (.25) $t_{(130)}$=11.07, p<.001. Again, the SAT passages yielded significantly lower scores.

The R^2's in Table 8 are based on the correlations between the recruits' responses in the ASVABpc and SAT conditions. They indicate that the better readers had more similarity in their multiple choice, as well as their Cognitive Structure responses, than the poorer readers, regardless of the text type. Since the better readers are likely better test takers, this stronger correspondence between the Cognitive Structure scores and the multiple-choice scores could be attributed to test-taking ability. The weak correspondence between the Cognitive Structure and

Table 8

Means and R^2's for the Cognitive Structure Scores and the Multiple-choice Scores for High and Low Readers

High Readers

Cognitive Structure Scores $R^2 = .22$

Variables	Means	SD	
ASVABpc	.56	.20	
SAT	.39	.24	($t_{(117)} = 8.38$, $p<.001$)

Multiple-Choice Scores $R^2 = .23$

Variables	Means	SD	
ASVABpc	66.58	21.47	
SAT	35.01	15.24	($t_{(137)} = 18.99$, $p<.001$)

Low Readers

Cognitive Structure Scores $R^2 = .07$

Variables	Means	SD	
ASVABpc	.44	.18	
SAT	.25	.14	($t_{(130)} = 11.07$, $p<.001$)

Multiple-Choice Scores $R^2 = .14$

Variables	Means	SD	
ASVABpc	53.21	19.06	
SAT	25.11	8.85	($t_{(151)} = 19.47$, $p<.001$)

multiple-choice scores in the SAT condition for the low ability readers may have been due to the fact that three of

the eight multiple-choice averages in this condition were at chance levels, i.e., any correlation involving a random variable will be random.

In conclusion, the correlations from Table 5 indicate that the two types of test have very similar construct validity. They are correlated with many of the same measures, and the recruits with high general knowledge verbal scores had good correspondence between their Cognitive Structure scores and their multiple-choice scores. The fact that the tests are less correlated, and therefore measure less similar constructs in the ASVABpc condition remains an unsolved issue.

<u>Interpreting the Factors.</u> The factor groupings of the subtests were interpreted as a vocational factor, a reading factor, and a math factor. The ASVAB subtests tests were designed to measure vocational aptitude, and it is therefore not surprising that the vocational factor emerged as the most general factor. However, the emergence of these distinct factors was encouraging news for advocates of the ASVAB. This test has been criticized in the past for failure to discriminate different abilities of the test takers (Murphy, 1984). The factor analyses demonstrated that the subtest scores do provide information about distinct abilities.

When the Cognitive Structure scores and the multiple-choice scores were included, the reading factor became a more general factor, and the Technical School Scores moved from the math factor to load higher on this first factor. It

should be noted that the Technical School scores still loaded relatively high on the math factor (.42 on Factor 1, and .39 on Factor 3).

The fact that the Cognitive Structure scores and the multiple-choice scores loaded at such high levels on the reading factor was good evidence for the relative construct validity of these two tests.

In conclusion, both the Cognitive Structure and the multiple-choice tests had some construct validity as tests of reading comprehension. They systematically varied with passage difficulty (both tests had lower scores for the SAT than for the ASVABpc) and general verbal ability (the lower ability readers scored lower on both tests). Nunnally's suggestion that construct validity can be investigated by examining variation in test scores that can be predicted by other established measures was proven useful here. Both tests loaded very high on the verbal or reading factor (multiple choice=.68; Cognitive Structure=.66) and low on the math (multiple choice=.29; Cognitive Structure=.24) and vocational factors (multiple choice=.26; Cognitive Structure=.21). The subsequent regression analyses for the predictive validity of these tests will help to triangulate on the actual constructs measured, but the statistical analyses presented here suggest that both tests were equivalent in their measurement of reading comprehension.

Predictive Validity Results and Discussion

External Criterion. Table 9 shows the regression models used to predict the Technical School final course grades for the recruits, along with supporting statistics. Unless otherwise noted, the entry method for the variables was a forced entry method. The corresponding reported Beta coefficients (in parentheses after the variable name in the text) were standardized to compensate for the fact that the multiple-choice scores and the Cognitive Structure scores were measured on different scales. Although these Beta coefficients do not give absolute information regarding the influence of each variable (due to correlations among the independent variables and between the predictor and the criterion variables), they were interpreted as rough indices of the contribution of each variable. None of the regression models resulted in variables with tolerance indices lower than .50, variance inflation factors over 2.0, and collinearity diagnostics with condition indices above 10.00.

When all passages were combined, the variables significantly predicted the Technical School scores, and explained about 19% of the total variance in those scores.

For the ASVABpc passages, the model significantly predicted the Technical School scores and explained approximately 13% of the total variance in scores. While both variables significantly contributed to the model, the Cognitive Structure scores (.27) had slightly more influence

Table 9

Predicting Technical School Scores

ASVABpc Passages

Model: TECHSCHOOL = CS + MC

$R^2 = .13$ $F_{(2,219)} = 16.74$, $p < .001$

Variable	Beta	t	p
CS	.2737	4.08	.001
MC	.1628	2.42	.016

SAT Passages

$R^2 = .14$ $F_{(2,222)} = 18.30$, $p < .001$

Variable	Beta	t	p
CS	.1688	2.06	.045
MC	.2417	2.89	.004

All Passages Combined

$R^2 = .19$ $F_{(2,187)} = 21.59$, $p < .001$

Variable	Beta	t	p
CS	.2583	2.97	.003
MC	.2179	2.51	.013

(i.e., explained more of the variance) than the multiple-choice scores (.16). The correlations among these variables also supported this claim with the Cognitive Structure scores explaining about 5% more of the variance in Technical School scores than the multiple-choice scores (see Table 5).

Considering the SAT passages only, the combination of Cognitive Structure scores and multiple-choice scores significantly predicted the Technical School scores. Again, the model explained about 14% of the overall variance in scores. Both variables significantly contributed to the model, but with these SAT passages, the multiple-choice scores (.24) explained slightly more of the variance than the Cognitive Structure scores (.17). However, the actual correlations in Table 6 suggested that this difference in influence was smaller than in the ASVABpc passages, with the multiple-choice scores explaining only 2% more of the variance.

Again, Nunnally's suggestions were followed by splitting the data at the median for General Knowledge Verbal (62.5), and the identical analyses were performed, (see Table 10). For the lower ability readers, the multiple-choice scores did not significantly predict the Technical School scores, while the Cognitive Structure scores did. For the higher ability readers, neither test alone significantly predicted Technical School scores (although the combination of the two variables did).

Table 10

Predicting Technical School Scores as a Function of General Verbal Ability

Low General Verbal Ability

$R^2 = .12$ $F_{(2,96)} = 6.31$, $p = .003$

Variable	Beta	t	p
CS	.2724	2.64	.01
MC	.1272	1.23	NS

High General Verbal Ability

$R^2 = .14$ $F_{(2,88)} = 7.19$, $p = .001$

Variable	Beta	t	p
CS	.1403	1.00	NS
MC	.2618	1.86	NS

To further investigate the locus of these effects, the split data were also analyzed by passage type (ASVABpc vs. SAT; see Table 11). For the ASVABpc passages and lower ability readers, the Cognitive Structure score significantly contributed to the model while the multiple-choice score did not. This was true for the higher ability readers as well.

Similarly, for the SAT passages and lower ability readers, the Cognitive Structure scores significantly contributed to the model while the multiple-choice scores did not. However, for the SAT passages and the higher ability readers, the opposite was true. That is, the multiple-choice scores significantly contributed while the Cognitive Structure scores did not.

For the passages intended to measure reading comprehension for this population (ASVABpc), the Cognitive Structure scores were a better predictor of Technical School scores than the multiple-choice scores. The median general verbal knowledge split revealed that the locus of this effect resided with the lower ability readers, that is, neither test was predictive of the higher ability readers. (The fact that the model was significant in this case could have been due to multicollinearity among the variables, but the correlation between the Cognitive Structure scores and the multiple-choice scores, while high ($r = .72$, $p<.001$), was not singular.) The Cognitive Structure scores were predictive of the Technical School scores of lower ability readers, and the multiple-choice scores were not. However, the combination

Table 11

Predicting Technical School Scores as a Function of Passage Type and General Verbal Ability (GVA)

Passage Type		General Verbal Ability	
ASVABpc		Low General Verbal Ability	
$R^2 = .09$ $F_{(2,114)} = 5.86$, $p = .004$			
Variable	Beta	t	p
CS	.2466	2.67	.009
MC	.1262	1.37	NS
ASVABpc		High General Verbal Ability	
$R^2 = .08$ $F_{(2,100)} = 4.86$, $p = .01$			
Variable	Beta	t	p
CS	.2338	2.30	.023
MC	.1216	1.20	NS

Table 11

Predicting Technical School Scores as a Function of Passage Type and General Verbal Ability (GVA)

Passage Type		General Verbal Ability	
SAT		Low General Verbal Ability	
$R^2 = .08$	$F_{(2,111)} = 4.628$, $p = .01$		
Variable	Beta	t	p
CS	.2048	2.12	.036
MC	.1311	1.36	NS
SAT		High General Verbal Ability	
$R^2 = .13$	$F_{(2,101)} = 7.79$, $p < .001$		
Variable	Beta	t	p
CS	.0623	.45	NS
MC	.3114	2.20	.03

of Cognitive Structure scores and multiple-choice scores did significantly predict performance for the higher ability readers, suggesting that each test measured some separate aspects of reading comprehension (or test taking ability), that when combined, significantly predicted the Technical School scores.

The partitioning of the data into passage type revealed that the multiple-choice scores were not predictive for either low or high readers when the ASVABpc passages were used. For the SAT passages, which may have been too difficult for these readers, the Cognitive Structure scores predicted for the low ability readers, while the multiple-choice scores predicted for the higher ability readers.

In conclusion, the Cognitive Structure scores were better predictors of the Technical School scores than the multiple-choice scores, especially for the texts that are arguably more appropriate for these participants (i.e., the ASVABpc passages). Additionally, the Technical School scores loaded highest on the verbal factor, suggesting a strong verbal component for this score. Since the Cognitive Structure scores were a better predictor of these scores, more evidence of the validity of the Cognitive Structure Test as a test of reading comprehension was provided by these results.

Domain Specific Knowledge. Domain specific knowledge was measured by performance on the subtests of the ASVAB. In general, all of the passages could be considered to be about

science or technical knowledge (mechanical, physical, biological, chemical, economic, tactical). The ASVABpc passages covered mostly vocational or technical knowledge (e.g., Autoshop and Mechanical Comprehension), while the SAT passages were primarily concerned with General Science. Considering all passages combined, both the Cognitive Structure scores and the multiple-choice scores were predictive of ASVAB General Science scores (see Table 12).

For the ASVABpc passages, both the Cognitive Structure scores and the multiple-choice scores were predictive of General Science, Autoshop, Mechanical Comprehension, and Electrical Information. The beta coefficients were quite similar to each other.

Table 12 also contains the information relative to the SAT passages for which General Science reflects the only domain specific knowledge represented by the texts. Here, both tests were predictive, with the multiple-choice scores (.34) accounting for approximately 7% more of the variance than the Cognitive Structure scores (.22).

The criterion variables used in these analyses were subtests of the ASVAB, a test designed for this population. Therefore, it would be expected that the multiple-choice scores for the ASVABpc passages would be predictive of these domain specific tests. But the results presented here give no objective reason for choosing one test over another on the grounds of domain specific knowledge predictability. The Cognitive Structure scores were equivalently predictive of

Table 12

Predicting Domain Specific Knowledge with the Cognitive Structure and Multiple-Choice Scores.

All Passages

General Model

General Science = CS + MC

$R^2 = .31$ $F_{(2,244)} = 53.92$, $p<.001$

	Beta	t	p
CS	.20	2.73	.007
MC	.41	5.38	.000

SAT Passages

General Science = CS + MC

$R^2 = .26$ $F_{(2,289)} = 50.44$, $p<.001$

	Beta	t	p
CS	.2150	3.11	.002
MC	.3371	4.87	.001

ASVABpc Passages

General Science = CS + MC

$R^2 = .16$ $F_{(2,284)} = 27.56$, $p<.001$

	Beta	t	p
CS	.2234	3.82	.001
MC	.2613	4.42	.001

Table 12

Predicting Domain Specific Knowledge with the Cognitive Structure and Multiple-Choice Scores

ASVABpc Passages

Autoshop = CS + MC

R^2 = .09 $F_{(2,284)}$ = 14.61, p <.001

	Beta	t	p
CS	.1323	2.16	.031
MC	.2214	3.75	.001

Mechcomp = CS + MC

R^2 = .12 $F_{(2,284)}$ = 19.29, p <.001

	Beta	t	p
CS	.1636	2.63	.009
MC	.2527	4.19	.001

Electric Information

R^2 = .1245 $F_{(2,284)}$ = 20.20, p <.001

	Beta	t	p
CS	.2224	3.67	.001
MC	.2016	3.39	.001

this domain specific knowledge that the ASVAB tests were intended to measure. The correlations presented in Table 5 supported this notion. For example, for the ASVABpc passages, the correlation between the Cognitive Structure score and the General Science score was equivalent to the correlation between the multiple-choice score and the General Science score (.32). Similar coefficients were also found among the other measures (Autoshop, Electric Info, and Mechanical Comprehension).

General Knowledge. Three criterion variables were used to assess general knowledge and general ability. First, the Nelson Narens General Knowledge Verbal was used as a proxy for general knowledge and reading comprehension. Second, a composite of ASVAB subtests, called the Armed Forces Qualification Test (AFQT) was used as a measure of general ability [AFQT = 2 X (Paragraph Comprehension + Word Knowledge) + Arithmetic Reasoning + Math Knowledge]. Third, another composite was calculated that reflected general knowledge (General Knowledge Verbal + General Knowledge Spatial + Word Knowledge + Math Knowledge). This composite contained two external tests (Nelson Narens General Knowledge Verbal + General Knowledge Spatial) and two subtests of the ASVAB (Word Knowledge + Math Knowledge). Word Knowledge was chosen (as opposed to paragraph comprehension) because it loaded highest on the reading comprehension factor and Math Knowledge loaded highest on the math factor (see Table 7).

For General Knowledge Verbal, when all passages were combined, the model significantly predicted this variable with the Cognitive Structure scores (.27) and the multiple-choice scores (.32) explaining similar amounts of the variance (see Table 13).

The same was true for the AFQT criterion and the General Knowledge Composite (GKC). However, the influence of (variance explained by) the multiple-choice scores was higher for the AFQT and the General Knowledge Composite than for the General Knowledge criterion, while the influence of the Cognitive Structure scores was relatively constant for all of these criteria.

It was not surprising that the multiple-choice scores were predictive of general knowledge and ability, when the regressor was a version of the paragraph comprehension subtest and the criteria were composites of the subtests from the same battery (or different versions of the same battery). A high degree of similarity among these test questions would be expected. However, the Cognitive Structure scores contributed an equivalent amount to this predictability even when multiple-choice tests similar to the other regressor were used.

In conclusion, these results demonstrated that the Cognitive Structure Test was as predictive of general knowledge and ability as the multiple-choice tests.

Table 13

Predicting General Knowledge with the CS and MC Scores

General Knowledge Verbal

$R^2 = .30$

$F_{(2,246)} = 52.09$ $p < .001$

	Beta	t	p
CS	.2723	3.80	<.001
MC	.3248	4.53	<.001

AFQT

$R^2 = .33$

$F_{(2,244)} = 60.77$ $p < .001$

	Beta	t	p
CS	.2513	3.58	<.001
MC	.3777	5.38	<.001

General Knowledge Composite

$R^2 = .36$

$F_{(2,244)} = 69.70$ $p < .001$

	Beta	t	p
CS	.2533	3.70	<.001
MC	.4037	5.89	<.001

The Experts.

Table 14 contains the correlations between the Cognitive Structure responses of the two experts from each domain, as well as the percentage correct on the multiple choice questions for each passage. For the ASVABpc passages, the experts were significantly correlated on all passages. Seven of the eight SAT passages (all but Surfaces) resulted in significant correlations between the two experts' Cognitive Structure responses.

The mean correlation in the ASVABpc passages was .75, and for the SAT passages was .62 (df=6). The experts correctly answered, on average 93.8% of the ASVABpc questions, and 82.86% of the SAT questions.

The experts were in relatively good agreement on the relationships among the terms (the Cognitive Structure responses) as well as to the answers to the multiple-choice questions, regardless of the passage type. It is interesting to note that these experts did not perfectly agree with the "correct" multiple-choice answers. One expert refused to answer a question, stating that there was no "correct" answer presented in the alternatives, and another expert stated that all of the questions were bad. In fact, many of the experts lamented on the quality of the multiple-choice questions.

All of the experts were faculty members at a large University, except for some of the experts for the ASVABpc passages, who were professionals in Automechanics

Table 14

Experts' Correlations and Multiple-choice Responses

ASVABpc			SAT		
Passage	rCS	MC%	Passage	rCS	MC%
Alka	.88	100	Bact	.61	75
Auto	.88	100	Clim	.55	75
Infl	.88	100	Glac	.65	90
Ligh	.95	100	Jets	.81	90
Phys	.55	100	Newt	.88	70
Smok	.49	50	Pros	.48	75
Tact	.72	100	Radi	.78	90
Wood	.63	100	Surf	.23*	75

Note. rCS = the correlation between the two experts' cognitive structure responses. *Not significantly greater than zero.

and Electrical Engineering. However, it was the SAT experts who mostly disagreed about the multiple-choice questions. Furthermore, 50% of the experts who read the ASVABpc passages were correlated above .88 (with the other subject matter expert), while only 12.5% of the experts who read the SAT passages were correlated this high. The descriptive statistics from the recruit population were used to infer that the SAT passages were harder than the ASVABpc passages, but the responses of the experts suggested that the significant differences in performance of the recruits could have been due to question ambiguity rather than difficulty. Psychometrically, the SAT passages and questions used here might be expected to be much more reliable (and perhaps valid) than the ASVABpc passages and questions. Each SAT passage was accompanied by at least 3 questions, and usually 4 or 5, while the ASVABpc passages had only 1 question for each. The responses of the experts to the multiple-choice questions suggested that there may be more wrong with these SAT type questions than even the missing-passage data of Katz, et al. implied.

Graphically Representing the Structures

The Cognitive Structure Test elicits responses from the examinees that were very suitable to graphic representation. Distance information in these graphs was interpreted as information about the semantic organization of the concepts or terms. To illustrate this, two passages were selected, one from the ASVABpc (Car in the Curve) and one from the SAT

(Surfaces). Addtree and Extree solutions were calculated for the matrix of responses of one expert from each of these passages. The graphs for "Car in the curve" (Figure 5) and for "Surfaces" (Figure 6) were visually inspected in order to make claims about the knowledge representations of the experts. The accompanying passages and terms are in Appendix B.

Figure 5 depicts the Addtree and Extree solutions for one expert who read the "Car in the Curve" passage. This is a short passage about the forces involved in a car rounding a curve. The root of the tree in the graph was chosen to minimize the variance between it and the terms. Both graphs had two main branches, with the terms "law of inertia" and "car goes straight" on the top branch, indicating that for this expert, these ideas belong together more so than with the other terms or ideas. The other four terms are further subdivided into two branches, one with "centripetal force" and the other with "road curves" as a separate branch connected to the cluster "car curves" and "turn steering wheel". Starting with the right most branch, this structure was interpreted as corresponding to a reading of the passage in which the car curves (as a result of) turning the steering wheel (as a consequence of) the road curving (resulting in) centripetal force. The law of inertia (causes) the car to go straight.

The letters on the Extree graph indicate that those terms with the same letter were more alike than the line

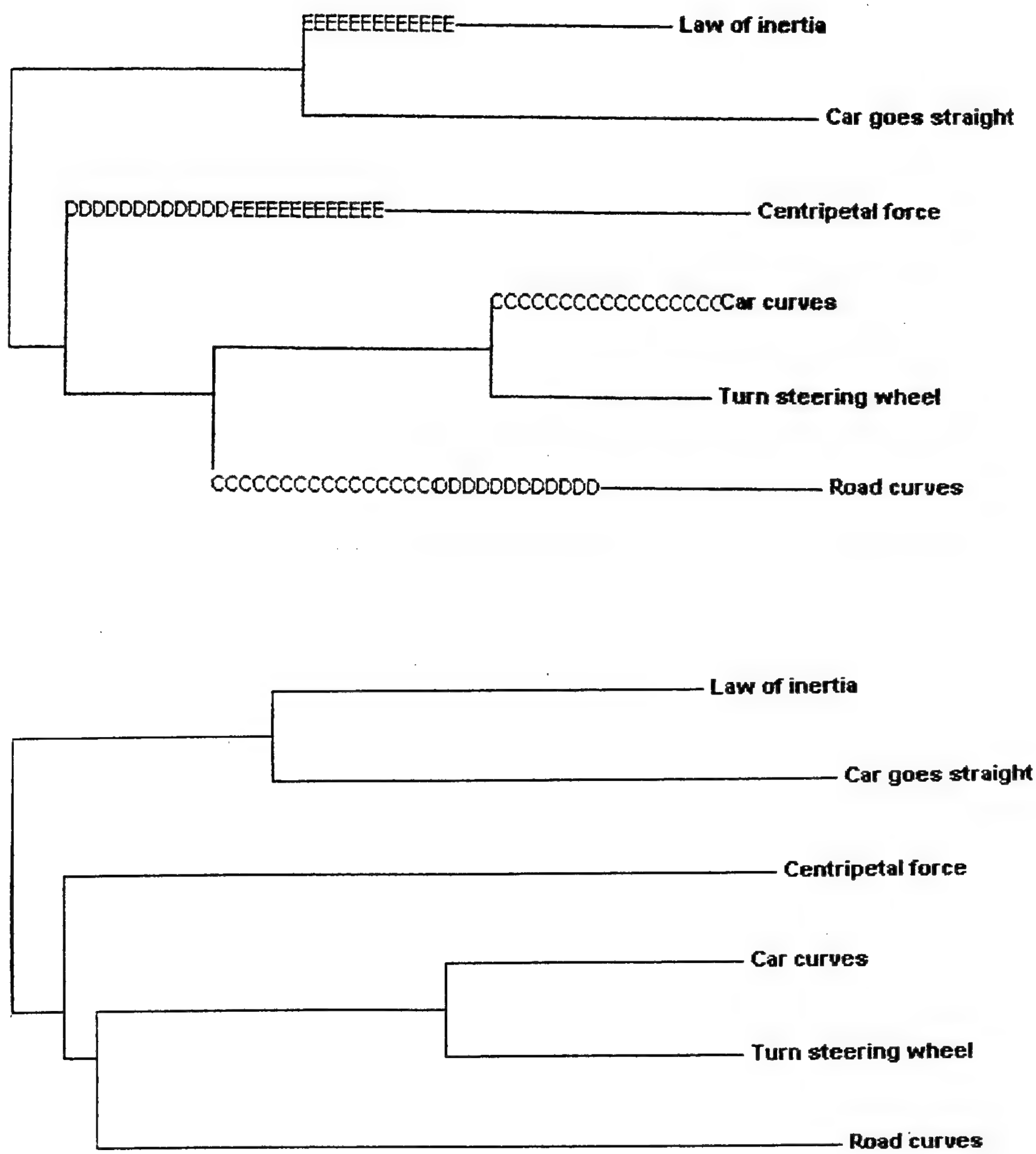
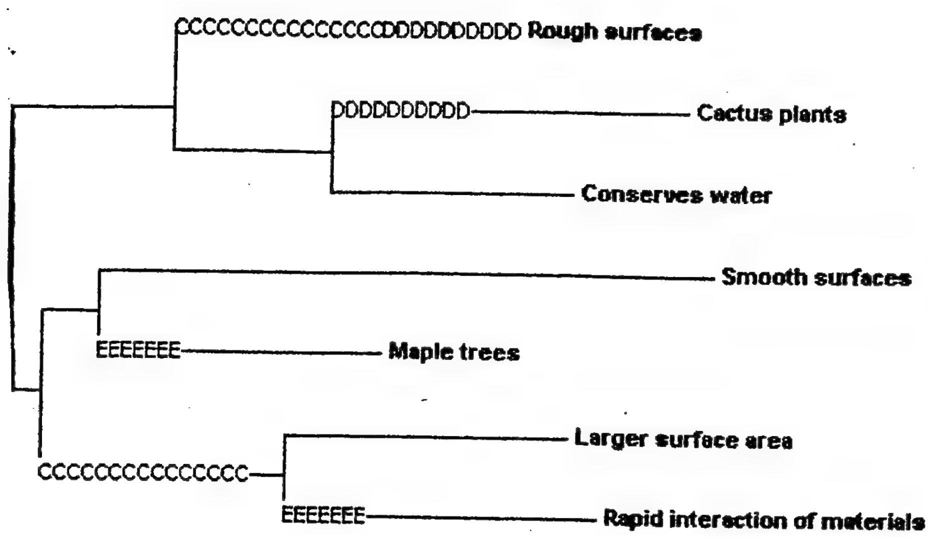


Figure 5. Addtree (bottom) and Extree (top) Graphs for Car in the Curve



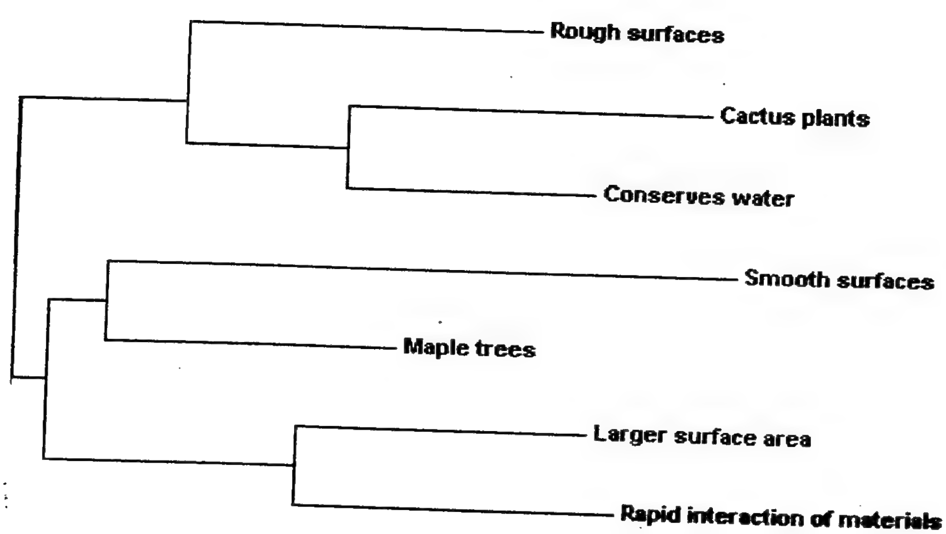


Figure 6. Addtree (bottom) and Extree (top) Graphs for Surfaces

groupings indicated. For this expert, "law of inertia" was related to "centripetal force" (they are both forces in the passage), "centripetal force" was related to "road curves" (this force is exerted in a curve), and "car curves" was related to "road curves" (these events are causally and temporally related).

Figure 6 was interpreted similarly. This passage described the nature of organic surfaces. This expert's responses were represented by two main branches. The top branch consisted of the terms ""cactus plants" and "conserves water" as a separate grouping, and "rough surfaces". The bottom branch was divided into two separate sub-branches, one with the terms "smooth surfaces" and "Maple trees", and the other with "larger surface area" and "Rapid interaction of materials". This corresponded with the passage which divided organic surfaces into rough and smooth, each with an example plant (cactus and maple tree, respectively), and an example property (conserves water, larger surface where materials interact rapidly).

The Extree graph indicated that the rough surfaces were related to the larger-surface-area-rapid-interaction-of-materials branch, and to the cactus plants, and that maple tree are related to rapid interaction of materials more so than indicated by the groupings alone.

In conclusion, these graphical representations provided information about the examinee's knowledge structures that was easily interpreted, and corresponded to the passages in

meaningful ways. Objective means of using these structure to diagnose readers' misconceptions should be empirically pursued.

General Discussion

Theoretical Considerations

It has been argued that the Cognitive Structure Test was superior to current measures of reading comprehension in that those measures were constructed and conducted atheoretically. The Cognitive Structure Test, on the other hand, was constructed to elicit responses that convey information more closely resembling what we know of reading comprehension, i.e., that it involves structuring a representation of the relationships among concepts, that it is dynamic management of working memory, and involves making inferences.

The Cognitive Structure Test more densely sampled the concepts in the domains, than the divergent concepts sampled in multiple-choice tests. For example, in the ASVABpc passage about alkalies (see Appendix B), the accompanying multiple-choice question asked the reader about the relationships between alkalies and four other terms (inert, radioactive, hydroxide, and water). Presumably, the reader assessed these relationships and decided which best answered the question. The Cognitive Structure test for this passage utilized 7 terms to assess 21 relationships. Similarly, in the longer SAT passage about bacteria, all of the multiple-choice questions (4) were questions about adhesion (a property of

bacteria surfaces). Nine terms from the text were used in the question alternatives, but the correct responses to the questions involved only knowledge of adhesion. In fact, the term "glycocalyx" appeared five times in this text, but was not mentioned at all in the questions. The Cognitive Structure test for this passage asked the reader for relatedness judgements for 10 different pairs of terms. This dense sampling of constructs and the ensuing relatedness judgments among them resulted in matrices of information that afford many graphical representations. These spatial representations have in the past provided useful information about the structure of knowledge (Jonassen, 1990; Britton & Sorrells, in press). Freedle and Kostin (1994) claimed that the construct validity of a test could be better demonstrated if it could be shown that the readers has constructed a coherent representation of the text. These graphical representations provided such a demonstration. Additionally, these structures have in the past provided reliable measures of learning that were useful for differentiating levels of achievement (Naveh-Benjamin, McKeachie, Lin, & Tucker, 1986).

Diekhoff's (1983) concern over developing testing methods that encourage students to consider the relationships among concepts was met with the Cognitive Structure Test's relationship judgments. This type of testing, Diekhoff contends, can be used to evaluate as well as promote structural knowledge.

Taking the Cognitive Structure Test also involved management of working memory, along with inference generation. To make these judgments, examinees presumably had to consider aspects of each term, holding these aspects in working memory until a judgment was reached. Working memory constraints and the similarity of the terms used imposed certain limitations on the Cognitive Structure Test. Informal observations and experience with this test attest to its difficulty. The novelty of the test procedure and of the representation of the results can provide motivation to offset these limitations. Additionally, inferences about the relationship among many of the terms must have been made since these inferences are by nature not explicit in the text. The multiple-choice tests can not make any of these claims.

The type of information provided by the Cognitive Structure Test is very suitable to connectionist analyses. The responses to the Cognitive Structure Test can be arranged in a matrix that represents the relatedness between all of the terms. The terms would be the labels for the columns and rows of this matrix, and the relatedness ratings for the pairs of terms would be the cell entries. This matrix of associations can be represented as a neural network with the nodes consisting of the terms, and the links between them determined by the relatedness ratings. Several algorithms exist (e.g., Kintsch, 1988) that can mathematically simulate spreading activation through these networks. The information

from multiple-choice tests of reading comprehension are not as yet suitable for these types of analyses.

Improving the Cognitive Structure Test

The Cognitive Structure Test could benefit from utilizing more of what theory has to offer. It is a wide contention that text difficulty, and subsequent difficulty in comprehending the text, is a function of the amount of argument overlap and required inferencing (e.g., Kintsch, 1988). The Cognitive Structure Test (as well as multiple-choice tests) could utilize texts that systematically vary in this overlap and inferencing. Since there are computer programs available to measure these variables, constructing texts that vary along these dimensions should provide no problem.

Another consideration is the motivation of participants who take the Cognitive Structure Test. The experts were tested individually and had the opportunity to ask questions about the Cognitive Structure Test. They were shown graphs like those in the Figures 4 and some of the experts adjusted their responses accordingly. The recruits were tested in a large group setting and did not have the same opportunities. The experts often lamented that they were having trouble responding to the test items. This consternation was met with a "do the best you can" reply. As more empirical evidence is provided for the use of the Cognitive Structure Test, better instructions need to be developed to provide examinees with

more information about the nature of the test and the meaning of their responses.

Unresolved Issues

This study demonstrated that prior knowledge can be measured with the Cognitive Structure Test with out biasing the measurement of post reading knowledge. The posttest Cognitive Structure measures were not systematically influenced by the administration of either pretest. However, the participants were not informed of the domain from which the terms from the pretests were taken, or for the general purpose of the pretest relationship judgments. Therefore, it can not be determined if the responses provided any information about the prior knowledge of the participants. More stringent controls need to be provided for measuring prior knowledge with the Cognitive Structure Test.

The working memory constraints also need to be considered when constructing Cognitive Structure tests. Too many terms will result in a test that is tiresome and by nature confusing (because of the similarity of the terms). The particular limit of terms and questions is yet to be determined.

More empirical work needs to be done using the graphical depictions of the structures. Independent judges could be used to increase the reliability of the diagnoses of the reader's misunderstandings, along with developing ways to address those misunderstandings.

The experts chosen in this study agreed reasonably well on the relationships among the terms. However, a larger group of experts could be used to answer important questions about the nature of the structure of the "true" referent. It would not be surprising to find that the average structure of several experts would be more stable than an average of two. Furthermore, factor analyses of larger groups of experts could provide the flexible referent used in this study, but the referent would have the advantage of being based on several experts' responses instead of just one.

The current status of the Cognitive Structure Test is that it performed equally well to the multiple-choice testing as a predictor of domain specific knowledge, general knowledge, and general ability, and that the Cognitive Structure Test was superior to these other tests in predicting the Technical School scores. In addition to these psychometric advantages, the Cognitive Structure Test has many theoretical advantages in that it elicits structures that afford analyses that are meaningful in the light of current models. The use of more densely sampled concepts from a domain allows a more thorough test of the examinee's knowledge. Diagnostic potentials have also been demonstrated, and the Cognitive Structure Test is easy to modify, administer, and score. These advantages of the Cognitive Structure Test over current multiple-choice measures of reading skill make the development, empirical investigation, and deployment of this test imperative.

Summary.

Current multiple-choice tests of reading comprehension (e.g., the SAT and the Armed Forces Vocational Aptitude Battery; ASVAB) have been criticized on the grounds that they are constructed without regard to theories or models of how a reader analyzes and understands a text. This criticism, and others (e.g., the ability to answer many reading comprehension multiple-choice questions without the reading passage) have prompted many to question the construct validity of these tests.

This study compared both the construct validity and the predictive validity of a new test, called the Cognitive Structure Test, with two examples of current multiple-choice comprehension tests (the SAT and the AFVAB). Results showed that the Cognitive Structure Test was at least equivalent, and in some ways superior to the multiple-choice tests in construct validity. Regression analyses were used to assess the predictive validity of both types of tests using criteria in three domains: an external domain criterion (final Techschool grade point averages for the recruits), domain specific knowledge criteria (vocational aptitude and general science tests), and general ability criteria (general verbal knowledge, general ability, and general knowledge composite tests). For passages that were intended for this population (i.e., the ASVAB passages), the Cognitive Structure Test was significantly predictive of the final Techschool scores, while the multiple-choice test was not. Both tests were

equally predictive of domain specific knowledge and general ability, even when the criteria were biased toward the multiple-choice test.

The experts' responses and methods for graphically representing the structures elicited through the Cognitive Structure Test were reviewed. The mean correlation between Expert 1 and Expert 2's Cognitive Structure responses for the ASVAB passages was .75, and for the SAT passages was .62. The experts correctly answered 93.8% of the ASVAB passage multiple-choice questions, but only 82.9% of the SAT passage multiple-choice questions.

The response patterns of two experts were compared using Addtree and Extree graphs. These graphical depictions or pictures were used to make inferences about the knowledge representations inferred from the spatial organization of the graphs.

It was concluded that the Cognitive Structure Test was superior to the ASVAB and SAT multiple-choice tests in several ways. Evidence of the Cognitive Structure Test's construct validity as a test of reading skill was provided by the correlational techniques (e.g., the high loading of the Cognitive Structure Test on the verbal factor of the tests' factor analysis solution) and the predictive advantages as well.

The Cognitive Structure Test assesses and can be used to graphically represent the structure of the examinee's responses, thus utilizing knowledge obtained from current

modeling and research on reading comprehension. The Cognitive Structure Test, by comparing the examinee's responses to the responses of two different experts has provided a more flexible referent than multiple-choice tests (which have only one answer for each question). Finally, the Cognitive Structure Test can sample more of a domain than a multiple-choice test, and therefore can provide more information about what an examinee does and does not understand.

Suggestions for improving the Cognitive Structure Test were made, along with a review of the unresolved issues. The results of this study suggest that Cognitive Structure Testing is an efficacious means of eliciting the types of responses from readers that afford statistical analyses and graphical representations that provide information about a reader's comprehension that is more in line with current theory than the types of responses provided by multiple-choice tests.

References

Acton, W.H., Johnson, P. J., & Goldsmith, T.E. (1994). Structural knowledge assessment: Comparisons of referent structures. Journal of Educational Psychology, 86, 303-311.

Bartlett, F.C. (1932). Remembering: A study in experimental and social psychology. Cambridge England: Cambridge University Press.

Britton, B.K. & Graesser, A. C. (Eds.) (1996). Models of understanding text. New Jersey: Lawrence Erlbaum Associates.

Britton, B.K. & Eisenhart, F.J. (1993). Expertise, text coherence, and constraint satisfaction: Effects on harmony and settling rate. Proceedings of the Cognitive Science Society, 15, 266-271.

Britton, B.K. & Sorrells, R.S. (in press). Learning connectionist networks and thinking about them. Discourse Processes.

Britton, B.K. & Tidwell, P. (1995). Cognitive structure testing: A computer system for diagnosis of expert/novice differences. In P.D. Nichols, S. Chipman, & R.E. Brennen (Eds.), Cognitively diagnostic assessment. Hillsdale, NJ: Erlbaum.

Britton, B.K., & Gulgoz, S. (1991). Using Kintsch's model to improve instructional texts: Effects of inference

calls on recall and cognitive structures. Journal of Educational Psychology, 8, 329-345.

Charnes, N. (1979). Components of skill in bridge. Canadian Journal of Psychology, 33, 1-50.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Chi, M.T., Glaser, R., & Farr, M.J. (Eds.) (1988). The nature of expertise. Hillsdale, NJ: Erlbaum.

Collins, A.M., & Quillian, M.R. (1969). retreival from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Corter, J.E. & Tversky, A. (1986). Extended similarity trees. Psychometria, 57, 429-451.

Dansereau, D.F., Brooks, L.W., Holley, C.D. & Collins, K.W. (1978). Learning strategies training: Effects of sequencing. Journal of Experimental Education, 51, 102-108.

de Jong, T. & Ferguson-Hessler, M. (1986). Cognitive structures of good and poor novice problem solvers in physics. Journal of Educational Psychology, 78, 279-288.

Diekhoff, G.M. (1983). Testing through relationship judgments. Journal of Educational Psychology, 75, 227-233.

Diekhoff, G.M., & Wigginton, P.K. (1989). Factors of success in a volunteer adult literacy program. Reading and Writing, 1, 153-162.

Drum, P.A., Calfee, R.C., & Cook, L.K. (1981). The effects of surface structure variables on performance in

reading comprehension tests. Reading Research Quarterly, 16, 486-514.

Fenker, R.M. (1975). The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures. Instructional Science, 4, 33-57.

Freedle, R. & Kostin, I. (1994). Can multiple-choice reading tests be construct valid? Psychological Science, 5, 107-110.

Gernsbacher, M. A. (1990). Language comprehension as structure building. Hillsdale NJ: Lawrence Erlbaum Associates.

Gernsbacher, M. A. (1996). In B.K. Britton & A. C. Graesser (Eds.) Models of understanding text. New Jersey: Lawrence Erlbaum Associates.

Geeslin, W. E. & Shavelson, R.J. (1975). An exploratory analysis of the representation of a mathematical structure in students' cognitive structures. American Educational Research Journal, 12, 21-39.

Goldman, S., Varma, S., & Cote, N. (1996). In B.K. Britton & A. C. Graesser (Eds.) Models of understanding text. New Jersey: Lawrence Erlbaum Associates.

Graesser, A.C. & Britton, B.K. (1996). Five metaphors for reading comprehension. In B.K. Britton & A. C. Graesser (Eds.) Models of understanding text. New Jersey: Lawrence Erlbaum Associates.

James, W. (1890). Principles of psychology. New York: Holt.

Johnson, P.E., Cox, D.L., & Curran, T.E. (1970). Psychological reality of physical concepts. Psychonomic Science, 19, 245-247.

Johnston, P. (1984). Prior knowledge and reading comprehension test bias. Reading Research Quarterly, Winter, 219-239.

Jonassen, D., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. Hillsdale, NJ.: Lawrence Erlbaum Associates

Just, M.A., & Carpenter, P. A. (1987). The psychology of reading and language comprehension. Newton, MA: Allyn and Bacon.

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99, 122-149.

Katz, S., Lautenschlager, G., Blackburn, A., & Harris, F. (1990). Answering reading comprehension items without passages on the SAT. Psychological Science, 1, 122-127.

Katz, S. & Lautenschlager, G. (1994). Answering reading comprehension items without passages on the SAT-1, the ACT, and the GRE. Educational Assessment, 2, 295-308.

Katz, S. & Lautenschlager, G. (1995). The SAT reading task in question: A reply to Freedle and Kostin. Psychological Science, 6, 126-127.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95, 163-182.

Kruskal, J. & Wish, M. (1978). Multidimensional scaling. Sage University Press: Beverly Hills.

Lavin, D. (1965) The prediction of academic performance: A theoretical analysis and review of research. New York. Russell Sage.

Medina-Diaz, M. (1993). Analysis of cognitive structure using the linear logic test model and quadratic assignment. Applied Psychological Measurement, 17, 117-130.

Murphy, K. (1984). Armed service vocational aptitude battery. Test Critiques, 61-69.

Naveh-Benjamin, M., McKeachie, W., Lin, Y., & Tucker, D. (1986). Inferring students' cognitive structures and their development using the "ordered tree technique." Journal of Educational Psychology, 78, 130-140.

Nunnally (1978). Psychometric theory. New York: McGraw Hill.

Phillips, D.C. (1983). On describing a student's cognitive structure. Educational Psychology, 63, 225-234.

Preece, P.F. (1978). Exploration of semantic space: Review of research on the organization of scientific concepts in semantic memory. Science Education, 63, 547-562.

Pyrczak, F. (1972). Objective evaluation of the quality of multiple-choice test items designed to measure

comprehension of reading passages. Reading Research Quarterly, 8, 62-71.

Quillian, M.R. (1968). Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge MA: MIT Press.

Reif, F. & Heller, J.I. (1982) Knowledge structure and problem solving in physics. Educational Psychologist, 17, 102-127.

Reitman, J.S. & Rueter, H.H. (1980).Organization revealed by recall orders and confirmed by pauses. Cognitive Psychology, 12, 554-581.

Roth, W., Gabel, D., Brown, L., & Rice, D. (1992). A combined method of INDSCAL and cognitive mapping for describing changes in students' cognitive structures. Education and Psychological Measurement, 52, 769-779.

Rumelhart, D.E. (1980). On evaluating story grammars. Cognitive Science, 4, 313-136.

Rumelhart, D.E. & McClelland, J.L. (1986). Parallel distributed processing. Cambridge, MA: MIT Press.

Sattath, S. & Tversky, A. (1977). Additive similarity trees. Psychometria, 42, 319-345.

Shavelson, R.J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Experimental Psychology, 63, 225-234.

Shavelson, R.J. (1974). Some methods for examining content structure and cognitive structure in instruction. Educational Psychologist, 11, 110-122.

Shavelson, R.J. (1983). On quagmires, philosophical and otherwise: A reply to Phillips. Educational Psychologist, 18, 81-87.

Stephens, D. & Britton, B.K. (1992). Cognitive psychologists' cognitive structures: Stable expert differences and novice changes. Unpublished manuscript.

Schvaneveldt, R. (Ed.) (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Stanners, L.T. & Brown, R.F. (1982). Conceptual interrelationships based on learning in introductory psychology. Teaching of Psychology, 9, 74-77.

Tidwell, P. (1992). Using cognitive structure measures to revise text: Diagnosis and repair of readers mis- and missing conceptions. Unpublished Doctoral Dissertation, University of Georgia.

Tuinmann, J.J. (1973). Inspection of passages as a function of passage dependency of the test items. Journal of Reading Behavior, 5, 186-191.

Wainer, H. & Kaye, K. (1974). Multidimensional scaling of concept learning in an introductory course. Journal of Educational Psychology, 66, 591-598.

Weinberg, S. & Menil, V. (1993). The recovery of structure in linear and ordinal data: INDSCAL vs. ALSCAL. Multivariate Behavioral Research, 28, 215-233.

van den Bergh. (1990). On the construct validity of multiple-choice items for reading comprehension. Applied Psychological Measurement, 14, 1-12.

van Dijk, T.A. & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic Press.

Appendix A

The Experts

ASVABpc Passages

Alkali

1. Dr. Richard Hill. PhD Harvard, 1954. Chemistry. UGA Professor Emeritus; 1993.

2. Dr. Robert Phillips. Associate Professor, UGA.

Automobile

1. Bill Fox. UGA Automotive Center.

2. Byron Smith. Athens Technical Institute Automotive Dept. Instructor.

Car in the curve

1. Dr. David Cohen. Associate Professor. Physics, UGA.

2. Dr. Roger Dean. Associate Professor, UGA. Animal and Dairy Science Dept. (BS & MS Physics)

Inflation

1. Paul Bingham. Stockbroker. A.G. Edwards and Son.

2. Dwight Lee. Professor Economics, UGA.

Lightning

1. John Rivers. Professor Emeritus, Physics, UGA.

2. Richard Sorrells. Electronics Technician. Private Company.

Smoke

1. Roy Arnold. AutoTech. Campus Transit.

2. Richard Payne. Auto Mechanic. Auto Center, UGA.

Tactical

1. Colonel Lloyd. UGA ROTC Commander. Former fighter pilot.

2. Tony Cushenberry. Former bomber pilot. Current Atheletic Dept. UGA.

Wood

1. Dr. Julian Beckwith. Associate Professor. Forestry Dept.

2. Dr. Garret Van Wicklen. Assistant Professor. Dept of Agriculture Engineering, UGA. PhD. Cornell, 1982.

SAT Passages

Bacteria

1. Dr. William Wade. Associate Professor, UGA. Pharmacy.

2. Dr. Cathy White. Assistant Professor, Pharmaceutics, UGA.

Climate

1. Dr. Robert Wyatt. Professor, Institute of Ecology, UGA.

2. Dr. Vernon Meentemeyer. Climatology Research Laboratory, Professor Geology, UGA.

Glaciers

1.Dr. David Dallmeyer. Professor, Geology and Earth Sciences, UGA.

2. Dr. Gilles Allard. Professor Geology, UGA.

Jetstream

1. Dr. Thomas Mote. Assistant Professor Geography, Climatology Research Laboratory, UGA.

2. Dr. Vernon Meentemeyer. Climatology Research Laboratory, Professor Geography, UGA.

Prostagulins

1. Dr. Honigberg. Professor Medical Chemistry, College of Pharmacy, UGA.

2. Dr. Ben Iturrian. Associate Professor, Pharmacology, UGA.

Radio

1. Dr. C.R. Johnson. Associate Professor. Physics, UGA.

2. Dr. David Cohen. Associate Professor. Physics, UGA.

Surfaces

Dr. Scott Martin. Associate Professor, Animal and Dairy Science, UGA.

2. Dr. James Price. Professor, Pharmacy, UGA.

Theories

1. Dr. Richard Meltzer. Professor, Physics, UGA.

2. Dr. Scott Kleiner. Professor, Philosophy, UGA.

Appendix B

Terms, Texts, and Multiple-Choice Questions

ASVABpc Texts

ALKALIES

Cognitive Structure	Prior Knowledge
Alkali	Mercury
Reactive	Hg
Inert	Sodium
Kerosene	Na
Hydroxide	Gold
Sodium	Au
Potash	

The term "alkali" is applied to a compound with definite basic properties: one that combines with water to liberate hydrogen gas and forms an hydroxide. Common hydroxides include lye (from sodium) and potash (from potassium). The alkali metals are among the most reactive known. This poses problems for their handling, and they must be stored in kerosene or some other inert substance. Sodium and potassium are the most common of the metals. Lithium is less common, but found in most rocks. Rubidium and cesium are rare, and francium is known only for its radioactivity. Alkali metals

are important physiologically in water retention, and they have important chemical and industrial applications.

A substance can be called alkaline if it:

a. is inert

b. is radioactive

c. forms an hydroxide

d. retains water

AUTOMOBILES

Cognitive Structure	Prior Knowledge Terms
Gasoline	Clutch
Alcohol	Torque-converter
Engine-compatibility	Propeller-shaft
Carbon-monoxide-producer	Universal-joints
Overheating-engine	Pinion
	Ring-gear
	Final-drive

Engineers thought it absurd that the average automobile engine should be able to make efficient use of only a fraction of its fuel. They could not fathom why owners of cars tolerated engines that are prone to overheat, that spew carbon monoxide into the air, that are dependent on a complicated distributor head, and that cannot use different kinds of fuel. Alcohol as an alternative fuel is already a

practical reality. The big question is whether the present automobile engine can be made compatible at a low cost.

The problem with alcohol for powering automobiles is that it:

a. may not reduce air pollution

b. boils at a very low temperature

c. increases atmospheric humidity

d. must be made cost efficient

INFLATION

Cognitive Structure	Prior Knowledge
Cause-of-inflation	Macro-economics
Solution-of-inflation	Total-production
Print-money-faster	Micro-economics
Print-money-slower	Individual-producers
Governments	Laissez-faire
	Capitalism

Substantial inflation is a monetary phenomenon, almost always rising from a more rapid increase in the quantity of money than the output of goods and services. Of course, the reasons for the increase may be various. It takes time--measured in years--for inflation to develop; it takes time for inflation to be cured, and there is only one fundamental cure. The rate of increase in the quantity of money must be curtailed. In today's world, governments determine--or can determine--the quantity of money.

According to the paragraph, which of the following statements is true?

a. There is only one cause of inflation

b. Inflation can be cured with short term action

c. An increase in the production of goods and services will cure inflation.

d. Government action is the key to halting inflation.

LIGHTNING

Cognitive Structure	Prior Knowledge
Closeness-of-land-to-clouds	Electromotive-force
Chances-of-lightning-striking	Potential-drop
Objects-above-others	Potential-difference
Objects-below-others	Ampere
Bush-in-forest	Coulomb
	Positive-to-negative
	Negative-to-positive

Lightning is a gigantic spark, a tremendous release of energy between earth and cloud. The shorter the gap between earth and cloud, the greater the chance of discharge. Thus lightning tends to favor objects that thrust above the surrounding terrain. This might mean you sitting in a boat or the loan tree on the golf course.

Lightning is described as:

a. man made energy

b. a bolt from heaven

c. a release of electrical energy

d. a poorly understood phenomenon.

CAR IN THE CURVE

Cognitive Structure	Prior Knowledge
Law-of-inertia	Study-of-mechanics
Centripetal-force	Motion
Car-curves	Statics
Car-goes-straight	Study-of-thermodynamics
Turn-steering-wheel	Heat
Road-curves	Energy

When a car rounds a curve, the laws of physics play a part in how turning steering wheel makes the car react. The law of inertia tends to keep the vehicle moving in a straight line. Centripetal force, resulting from friction of the tires against the road, moves the car in the direction of the turned front wheels. At a critical speed, inertia may overcome tire-and-road friction, sending the car into a straight path, out of the curve, and possibly off the road.

Entering a curve below the critical speed will cause the:

a. centripetal force to outweigh the inertial force

b. car to move in a straight path

c. fraction of the rites against the road to diminish

d. car to leave the road or overturn.

SMOKE

Cognitive Structure	Prior Knowledge
Black-smoke	Clutch
High-speeds	Torque-converter
Slow-speeds	Propeller-shaft
Reduce-gas-supply	Universal-joints
Install-sticky-spring	Pinion
	Ring-gear
	Final-drive

Black smoke coming from the muffler indicates that the gasoline-air mixture from the carburetor is too rich. Not only is fuel wasted, but cylinders can become fouled and carbonized. At slow speeds the gasoline supply should be cut back. At high speeds the repair needed might be to install a sticky auxiliary air spring or metering pin.

All but one of the following might be involved with a gasoline-air mixture that is too rich. Which one is the exception?

a. Black exhaust from the muffler.

b. Poor fuel economy.

c. Reduced gasoline supply.

d. Carbon deposits on cylinders.

TACTILE

Cognitive Structure	Prior Knowledge
Early-in-strategic-bombing	Fighter-bombers
Destroy-enemy-civilian-morale	Strategic-bombers
Strengthen-enemy-civilian-morale	Attack-planes
Later-in-strategic-bombing	Tactical-air
Attack-factories	Strikes-far-behind-enemy-lines

To the theorists of the Tactical School, strategic bombardment was first visualized only as a means of destroying enemy civilian morale. The feasibility of direct attack on enemy centers stopped short of outright populations bombing. It was pointed out the Japanese strikes against Chinese cities only strengthened morale. Later, attacks on industrial targets during daylight were favored. The economy of an industrialized nation might be disrupted by disabling just a small number of factories.

The Tactical School:

a. favored daylight bombing because most workers would be in factories then

b. opposed outright attack on civilian populations as inhumane

c. believed bombing industrial targets ineffectively concentrated airpower

d. realized that bombing industrial targets would have a more devastating effect on the enemy's war effort

WOOD

<u>Cognitive Structure</u>	<u>Prior Knowledge</u>
Log-when-first-thrown-on-fire	High-heat
Produces-significant-heat	Disorder-of-molecules
Absorbs-heat	Low-heat
Burns-at-once	Increased-entropy
Charcoal	Decreased-entropy
Flammable-gases	

Wood does not begin burning instantly when a log is tossed onto a fire. In fact, heat is robbed from the existing fire to get the new log burning and producing heat. Heat from the existing fire turns excess water in the new log to steam. Still absorbing heat, the wood begins to break down into flammable gases and charcoal. In the third stage these gases begin to burn and produce some heat. In the fourth and final stage, the charcoal residue burns and gives off significant heat, the point at which you begin to benefit from a wood fire.

Which statement is correct based on the paragraph?

a. Two of the stages listed produce heat and two of the stages consume heat

b. Flammable gases are the major producers of heat in a wood fire.

c. Material other than wood will make the hottest fires.

d. Burning wood that is holding moisture increases the heat produced because it prolongs the fourth stage.

e. Steam is the hottest gas to escape from the fire.

BACTERIA

Cognitive Structure	Prior Knowledge
Coated-bacteria	Halogen
Uncoated-bacteria	Iodine
Glycocalyx	Nitrogen
Pure-cultures	Ammonia
Pure-cultures	Phosphorous
Natural-environments	ATP

Bacteria stick, often with exquisite specificity, to surfaces ranging from the human tooth or lung to a rock submerged in a stream. They do so by means of a mass of tangled fibers of polysaccharides, or branching sugar molecules, that extend from the bacterial surface an form a feltlike glycocalyx surrounding an individual cell or a colony of cells. The adhesion mediated by the glycocalyx explains particular locations of bacteria in most natural environments; more specifically, it is a major determinant in

the initiation of bacterial diseases ranging from dental cares to pneumonia.

These major--and with the benefit of hindsight, obvious--facts about the bacterial cell surface have become known only within the past decade. Ironically, the main reason for the later discovery of the bacterial glycocalyx was the long reliance of microbiologists on an otherwise eminently effective investigative system: the pure laboratory culture of an individual bacterial strain. To generate and maintain a glycocalyx, a bacterial cell must expend energy, and in the protected environment of a pure culture, the glycocalyx is a metabolically expensive luxury conferring no selective advantage. Cells that fabricate these elaborate coatings are usually eliminated from pure cultures by uncoated mutants that can devote more of their energy budget to proliferation. It is these uncoated mutants that microbiologists have usually studied.

Which of the following can be inferred from the passage about bacteria that expend energy to produce adhesive capacity in a laboratory culture?

a. They have a limited advantage over those that do not.

b. They are likely to proliferate more rapidly than uncoated bacteria.

c. They waste resources that would enhance their chances for survival.

d. They draw energy away from neighboring uncoated bacteria.

e. They are more resistant to disease than are uncoated bacteria.

The passage indicated that the failure of microbiologists to recognize bacterial adhesive properties is best described as

a. an inevitable results of studying laboratory cultures

b. a result of careless laboratory techniques

c. an astounding lack of scientific judgement

d. a confused reaction to previous overemphasis on bacterial mutations

e. a result of an inability to identify all the surfaces adhered to by bacteria

Which of the following statements concerning the specificity of bacterial adhesions is (are) supported by the passage?

I. It is indispensable in bacterial proliferation

II. It is less evident in laboratory cultures

III. It is unimportant in most natural cultures

a. I only b. II only c. III only d. I and II only

e. II and III only

The primary purpose of the passage is to

a. explain a previous gap in the knowledge of bacterial surface function

b. describe the various surfaces to which bacteria can adhere

c. call for reevaluation of current microbiological beliefs

d. demonstrate how bacterial disease can be initiated by adhesion

e. explore the possibility of making environmental bacteria nonadhesive

CLIMATE

Cognitive Structure	Prior Knowledge
Day-length	Temperature
Photo-periodism	Air-pressure
Temperature	Humidity
Precipitation	Barometer
Causes-diapause	Tornado
Causes-flowering	Precipitation

In any part of the world where there are pronounced seasonal changes in climate, organisms appear and disappear at particular times with uncanny precision. Biologists have long wondered how they do it, and the answer to this question is even now not entirely clear.

As the Earth spins throughout its seasonal cycle, several environmental parameters change on an annual basis. Temperatures fluctuate, periods of high and low precipitation alternate, and day length increases and decreases. Like temperature, the time and amount of precipitation in most

parts of the world is highly unpredictable on a week-to-week basis, even thought seasonal averages my not fluctuate significantly. Unlike temperature and precipitation, however, day length repeats itself with monotonous precision year in and year out; it is not particularly surprising, therefore, that many organisms respond to this parameter.

Response to day length, or photoperiodism, among most groups of higher organisms is well known. However, the most detailed studies have been done with flowering plants and insects--particularly because they are readily handled in large numbers under laboratory conditions, and partially because many of them have sufficiently short life cycles that their response to photoperiod is quickly evident and hence accessible to experimentation. In insects the most carefully studies phenomenon is entry into diapause, a dormant state that occurs in different insects at very different developmental stages, from the egg through adulthood. In flowering plants the phenomenon subjected to closest analysis has been the transformation from vegetative to reproductive growth--the initiation and development of flowers--although it has been demonstrated that other phenomena, such as the onset of autumn coloration and entrance into winter dormancy, are also responses to day length.

The passage is best described as which of the following?

a. A response to a controversial theory of ecological relationships.

b. An explanation of a newly discovered cycle of growth.

c. An account of climatic events in different parts of the world.

d. A list of observations made during the course of one year.

e. An inquiry into possible causes of certain biological phenomena.

The author cites all of the following as responses to photoperiodism <u>except</u>:

a. insect entry into diapause

b. development of flowers

c. short life cycles of plants and insects

d. onset of winter dormancy in plants

e. start of autumn coloration in leaves

According to the passage, the phenomenon of diapause is accurately characterized as which of the following?

I. A state of inactivity

II. A response to a change in the number of daylight hours

III. An effect produced primarily through artificial laboratory conditions

a. I only b. II only c. III only d. I and II only e. I, II, and III

GLACIERS

Cognitive Structure	Prior Knowledge
Plucking	Glacier
Berschrund	Iceberg
Cirque	Foot-of-glacier
Abrasion	Moraine
Causes-U-shaped-valleys	Gravity
Glaciers	Buoyancy
Streams	

Valley glaciers are powerful agents of erosion, which they bring about primarily by plucking and abrasion. Plucking takes place mainly near to "head" (the back wall) of the glacier where the ice freezes to the rock wall and pulls off slabs as the glacier moves forward. Usually a large crack or crevasse, called a bergschrund, develops between the ice and the wall of rock as a result of the tension at the head of the glacier. The bergschrund rarely becomes much wider than 10-15 feet, thought, because each winter it is filled with snow that in turn becomes part of the glacier, freezes to the rock wall, and continues to pluck rock as the glacier slowly moves downhill. The continuous plucking action of the glacier results in the formation of a deep amphitheater called a cirque.

Abrasion is caused by rock fragments that are frozen into the bottom and sides of the glacier. These fragments act as a huge file or rasp as the ice moves down its course. Valleys eroded by streams are V-shaped, but glaciers scour

and grind the sides and bottoms to form the characteristic U-shaped cross section found in practically all the valleys in glacier areas. Since the main glaciers are usually at least a half-mile thick, the results of glaciation can be seen far up the sides of the large valleys.

The tributaries of glacial valleys are also peculiar in that they usually enter high above the floor of the main valley and thus are known as hanging valleys. The thicker the stream of ice, the more erosion it is capable of causing; consequently the main valley becomes greatly deepened, whereas the smaller glaciers in the tributary valley does not cut so rapidly, leaving its valley "hanging" high above the floor of the major valley. Fortunately, the ice has melted from innumerable glacial valleys, exposing many hanging valleys to view. Among the most spectacular sights in countless glacial valleys are the tumultuous waterfalls that leap from hanging valleys and drop their thundering ribbons of white to the placid waters in the valleys far below.

Even more conspicuous than the large U-shaped valleys and their hanging tributaries are the sharp-crested ridges that form the background of glacial mountain ranges. These features, known as aretes, owe their origins to glaciers. As long valley glaciers enlarge their cirques and cut farther in toward the center of the range, this area is finally reduced to a very narrow steep-sided ridge, the arete. In certain places three or more glaciers on the same ridge pluck their

way back toward a common point, leaving at their heads a conspicuous, sharp, pointed peak known as a horn.

Which of the following titles would be most appropriate for this passage?

a. Glaciers and their effects on land

b. The glacier as an agent of destruction

c. How and why glaciers advance and recede

d. The last remnants of the ice age

e. The surprising beauty of glaciers

According to the passage, hanging valleys are formed by

a. glaciers that develop before main glaciers do

b. the plucking action of main glaciers high on mountain sides

c. the slower action of glaciers smaller than the main glacier

d. the rapid melting of tributary glaciers before main glaciers melt

e. glaciers that cause erosion by plucking rather than abrasion

As it is used in the passage, the phrase "thundering ribbons of white" emphasizes which of the following?

a. the erosive power of waterfalls

b. the ever-changing landscape in glacial regions

c. the danger of waterfalls at high elevations

d. the various forms of water in glacial regions

e. the majestic beauty brought about by glaciers

It can be inferred from the passage that the most extensive erosion in glacial regions results from the

a. alternate freezing and thawing as seasons change

b. cracking that occurs during the formation of a bergschrund

c. cutting action of melting ice

d. plucking action of a tributary glacier

e. abrasion action of a main glacier

The author develops the passage by doing all of the following except

a. analyzing the process that causes glacial movement

b. defining terminology relating to glacial action

c. including a simile to explain glacial erosion

d. describing the results of glacial erosion to another type of erosion

JETSTREAM

Cognitive Structure	Prior Knowledge
Build-up-of-jetstream	Temperature
Temperature-inversion	Air-pressure
Vertical-mixing	Humidity
Earth-becomes-cloud	Barometer

Convection-turbulence	Tornado
Mechanical-turbulence	Precipitation

The general features of the low-level jetstream are now fairly well known. On the days that it occurs, it begins to build up in the late afternoon; it reaches its maximum in the middle of the night and decays in the early morning. At the peak of the jet, the winds in its core, between 800 and 2,000 feet up, can attain between 50 and 80 miles an hour, decreasing to 10 to 20 miles per hour between 3,000 and 4,000 feet and to zero at the ground.

Meteorologist generally agree that the diurnal cycle of heating and cooling of the earth plays a major part in the development of low-level jets. On a clear day, as the ground soaks up solar energy and grows warmer, it heats the layer of air immediately above it. If this layer gets hot enough, it begins to push its way up through the cooler air above it and a convection pattern is set up, with warm air rising over some parts of the surface and cooler air descending over others. This convection turbulence, as it is called, mixes the air at different levels. Also contributing to the mixing is mechanical turbulence, which depend on the strength of the wind and the roughness of the terrain. The interchange makes the pattern of wind speeds through the first few thousand feet more nearly uniform than it would otherwise be. Air moving closest to the surface is subject to maximum frictional drag from the ground. As a result the wind at the

lowest levels is slower than in the layers above it. Mixing partly offsets this effect; the ascending parcels of air carry up with them their lower speeds, while the descending cells bring down their higher speeds.

If the day has been calm as well as clear, with little mechanical turbulence, mixing falls off sharply as the sun goes down and the heating of the lower air decreases. The lowest air layers, still affected by surface drag but cut off now from the momentum supplied from above during the day, move more and more slowly. At the same time the upper layers are no longer sapped by contributions of momentum to the surface layers or slowed by injections of slower surface air. The winds aloft therefore speed up, and the jet begins to form.

The build-up of the jet is assisted by another nighttime weather phenomenon: temperature inversion. During the day the temperature of the first few thousand feet of atmosphere generally decreases with height. After the sun goes down, the ground begins to lose heat by radiation. If there is no blanket of clouds, the surface and the adjacent layers of air soon become cooler than the air above them. Here the temperature increases with height through the first thousand feet or so, and each succeeding parcel of air thought the layer is warmer and lighter than the one below it. This is a stable arrangement, which further damps out vertical mixing. On nights when the jet develops, the depth of the inversion layer increases during the hours just after

sunset, and the fastest winds are just above the top of the deepening inversion.

The primary purpose of the author apparently is to

a. explain meteorological terminology

b. illustrate what is meant by "convective turbulence"

c. explain the effect of the diurnal cycle

d. describe the formation of low-altitude jet streams

e. relate temperature inversion to jet stream formation

According to the passage, temperature inversion is important in the formation of the low-level jet stream because it

a. reduces mixing

b. occurs at night

c. cools the surface air

d. creates layers of air

e. reduces heat loss by radiation

On the basis of the passage, when does a low-level jet stream seem most likely to be at maximum intensity?

a. Morning

b. Noon

c. Midafternoon

d. Evening

e. Midnight

By mixing the air from different levels, mechanical turbulence and convective turbulence tend to

a. create winds particularly at the higher levels

b. slow the lower-level air

c. speed up the upper-level air

d. produce temperature inversion

e. equalize the momentum at the two levels

It can be inferred from the passage that the conditions most essential for the formation of a low-level jet stream is

a. a rise in the temperature of the higher levels or air

b. a minimum of vertical mixing of air

c. surface drag on the lowest air layers

d. mechanical turbulence during the day

e. a maximum of momentum transfer between air layers

PROSTAGLANDINS

Cognitive Structure	Prior Knowledge
Prostaglandin	Carbohydrates
Synthesis-of-prostaglandin	Lipids
Aspirin	Proteins
Prevents-heart-attacks	Carbon
Reduces-platelet-clumping	Hydrogen
Makes-platelets-clump-together	Nitrogen
	Sulfur

Prostaglandins are short-lived hormonelike substances made by most cells in the body after injury or shock. They are responsible for a number of physiological reactions.

Prostaglandins have been shown to influence blood pressure, muscle contraction, and blood coagulation and are involved in producing pain, fever, and inflammation. When released from platelets--minute discs in the blood---a prostaglandin derivative called thromboxane makes the platelets clump together and thus initiates clotting.

In 1971, John Vane, a British researcher, discovered that aspiring interferes with the synthesis of prostaglandins. Scientists now know that aspiring relieves pain by inactivating cyclooxygenase, an enzyme that aids initiating the synthesis of prostaglandins. When scientists realized that aspirin can also interfere with clotting, they began to wonder whether it could help prevent heart attacks and strokes, which are often caused by blood clots that block arteries in the chest and neck. Studies now indicate that low daily doses of aspirin can cut the risk of a second heart attack by about twenty percent and the risk of a second stroke by nearly half. It seems logical to assume that if the drug can prevent second heart attacks, it can also ward off an attack the first time around. Therefore, many doctors recommend an aspiring tablet every other day to inhibit excessive platelet clumping among people who have high blood pressures or other symptoms that increase the risk of attacks.

According to the passage, prostaglandins play a role in all of the following <u>except</u> the

a. clotting of blood

b. sensation of pain

c. contraction of muscles

d. manufacture of platelets

e. inflammation of tissue

The passage suggests that which of the following would be most likely to initiate the production of prostaglandins?

a. Taking an aspirin

b. Spraining an ankle

c. Climbing stairs

d. Flexing a muscle

e. Running a fever

It can be inferred from the passage that when the production of prostaglandins is impeded, which of the following occur(s)?

I. Blood coagulation is slowed

II. Pain is reduced

III. Inflammation increases

a. I only   b. II only   c. I and II only   d. II and III only

e. I, II, and III

It can be inferred from the passage that aspirin helps prevent heart attacks by

a. interfering with the production of thromboxane

b. lowering blood pressure

c. easing muscular contractions

d. initiation the production of cyclooxygenase

e. widening the arteries

RADIO

Cognitive Structure	Prior Knowledge
John-Fleming	Electromotive-force
Diode	Potential-drop
Lee-De-Forest	Potential-difference
Audion	Ampere
Triode	Coulomb
Radio	Positive-to-negative
Gas-filled-tubes	Negative-to-positive

In February 1902 Lee De Forest wrote, "I shall move all heaven and earth to put in at once a broad fundamental patent on telephony without wires by hertzian waves." As a graduate student at Yale, he had specialized in the study of electromagnetic waves, a phenomenon discovered by Heinrich Hertz in 1887. His subsequent work focused on the inadequacies of the Branly coherer, a device designed to detect these waves and a critical link in Marconi's system of wireless telegraphy. Despite research and invention that had spanned a number of years, De Forest's plans remained unrealized until a completely reliable detector was found to replace the Branly coherer.

The situation at last appeared, thanks to an observations Edison made in 1883 when he put a metal-wire electrode with a positive charge into his light bulb and found that electricity flowed from the flowing filament to

the new electrode across the space between them. Twenty years later John Fleming discovered that the "Edison effect" bulb could detection hertzian waves in a completely new way. This was the first radio vacuum tube, the diode.

De Forest began experimenting with a simple detector that contained a gas flame instead of an electric filament. In addition, he used a telephone receiver--with a battery of its own--and clearly heard the dot-dash wireless signals from a distant transmitter. He added these two items to Fleming's diode and later tried a zigzag grid or wire between the filament and a metal-late electrode to carry the incoming signal. De Forest named this first triode electrode to carry the incoming signal. De Forest named this first (patented in 1907) the "Audion" and much later called it "the grandaddy of all the vast progeny of electric tubes that have come into existence since."

But the progeny were not immediately forthcoming, partly because no one fully understood how the Audion really worked. Some gasses remained in the tube's partial vacuum. To De Forest, it seemed that the current could only flow from filament to plate through a transporting medium--the ionized gases. It was an obvious explanation--but incorrect. By 1912 it had become clear that a transporting medium was unnecessary. The filament emitted electrons--particles with negative charge--and the plate with its positive charge attracted them. High-vacuum tubes would be far more efficient than the "soft" gassy tubes used by De Forest until then.

De Forest also discovered that by feeding part of the output off his triode vacuum tube into its grid, he could cause a self-regenerating oscillation in the circuit. The signal from this circuit was far more powerful and effective than that of the crude transmitters then generally employed. When appropriately modified, this single invention was capable of transmitting, receiving, or amplifying radio signals and it was not surpassed until the invention of the transistor in 1947.

According to the passage, the first triode could best be described as

a. a rudimentary device that was later refined and improved

b. a simple structure that functioned as efficiently as does a transistor

c. a complex machine that ultimately proved too intricate to be practical

d. an ill-conceived design that was quickly superseded by the work of other inventors

e. a controversial apparatus that stunned scientists of the time

De Forest's statements quoted in the passage reveal which of the following attitudes?

a. Impatience and modesty

b. Idealism and detachment

c. Determination and pride

d. Fascination and awe

e. Hostility and defiance

According to the passage, De Forest's explanation of how his original Audion functioned was incorrect because he misunderstood the

a. role of electrons in ionized gases

b. means by which electrons travel in a vacuum tube

c. rate at which electricity is absorbed by the plate

d. composition of the partial vacuum inside the tube

e. necessity for both a filament and a plate

It can be inferred from the passage that De Forest's triode was capable of which of the following?

I. Intensification of the strength of the signal produced by a transmitter

II. Production of supplementary electrical power

III. Amplification of incoming signals in the receiver

a. II only b. III only c. I and II only d. I and III only

e. I, II, and III

Which of the following titles is most suitable for the passage?

a. Founders of Modern Electronics

b. Personalities of Early Radio

c. The Discovery of Hertzian Waves

d. How to Use the High-Vacuum Tube

e. A Pioneer in Radio Technology

<u>Cognitive Structure</u>	<u>Prior Knowledge</u>
Smooth-surface	Carbohydrates
Rough-surface	Lipids
Cactus-plants	Proteins
Maple-trees	Carbon
Larger-surface-area	Hydrogen
Rapid-interaction-of-materials	Nitrogen
Conserves-water	Sulfur

Both living and nonliving things interact with their environments across their surfaces. The nature of the surface greatly affects the kinds of interactions that can occur. As a rule, a smooth surface has less functional surface area per square meter than a rough or wrinkled one, and a larger ratio of surface area to volume will permit more rapid interaction of materials. Finely granulated sugar, for instance, dissolves more rapidly in a cup of coffee than a lump of sugar does.

During the course of evolution, modification of surface area has been an important adaptation or organisms to the environment. The wall of the human small intestine, for example, is layered with microvilli that provide greater surface area for the absorption of food. Similarly, in plant roots epidermal cells have extensions (root hairs) that increase a plant's ability to take in water and minerals from the soil.

Unlike the interchanges that occur at the surfaces of most nonliving substances (such as the sugar cube), exchanges of materials in living cells occur through cell membranes that are selectively permeable. Cell membranes regulate the movement of most substances in and out of the cell, allowing passage of some substances and blocking others.

Cell membranes, however, appear unable to regulate the movement of water. Over the millennia, organisms have evolved many strategies to conserve water. Cactus plants have adapted to their hot, dry, environment by developing thickened leaves with the relatively reduced surface area. In areas where the lack of water is not so sever, most plants have leaves with smaller volumes and larger surface areas. Broad-leaved plants such as beech and maple trees lose water rapidly through transpiration and thrive where moisture is abundant and temperatures are moderate.

This passage primarily discusses

a. the difference between the surfaces of living things and the surfaces of nonliving things

b. how the roughness of an object's surface affects its permeability

c. in what manner various living things have adapted to their environments

d. why cell membranes are unable to control the absorption and release of water

e. how the surface area of living and nonliving things affects their interaction with their environment

In lines 13-18, the author mentions microvilli in human intestines and roots hairs of plant roots as illustration of adaptation that

a. decrease the cells' dependence on water

b. increase the absorptive ability of the organism

c. retard the exchange of material between the cells and the environment

d. maintain the flexibility of the cellular walls

e. decrease the surface area of the cellular walls

According to information in the passage, "most substances" (lines 23-24) would include which of the following?

I. Food

II. Water

III. Minerals

a. II only b. III only c. I and II only d. I and III only e. I, II, and III

It can be inferred from the passage that the surface of which of the following is least conducive to rapid interaction of materials?

a. a prune

b. a sponge

c. a piece of cake

d. a balloon

e. a tree root

NEWT (THEO) TERMS

Cognitive Structure	Prior Knowledge
Common-sense	Darwin's-theory
Scientific-theories	Creationist-theory
Unnatural-theories	Lorenz's-theory
Supports-desirable-behavior	Einstein's-theory
Undermines-desirable-behavior	Heliocentric-theory
Instruments-of-indoctrination	Galileo's-theory

In the eighteenth and nineteenth centuries, Newton's mechanics not only came to be thought of as compatible with common sense but had even been identified with common-sense judgement. As a results, in twentieth-century physics, the theory of relativity and the quantum theory were regarded by many as incompatible with common sense. These theories were regarded as "absurd" or, at least, "unnatural" and were rejected as Francis Bacon had rejected the Copernican system. Looking at the historical record, we notice that the requirement of compatible with common sense and the rejection of "unnatural theories" have been advocated with a highly emotional undertone, and it is reasonable to raise the question: What was the source of heat in those fights against new and "absurd" theories? Surveying these battles, we easily find one common feature: the apprehension that a disagreement with common sense may deprive scientific theories of their value as incentives for desirable human behavior. In other

words, by becoming incompatible with common sense, scientific theories lose their fitness to support desirable attitudes in the domain of ethics, politics, and religion.

Examples are abundant from all periods of theory-building. According to an old theory that was prevalent in ancient Greece and was accepted by such thinkers as Plato and Aristotle, the sun, planets, and other celestial bodies were made of a material that was completely different from the materials of which our earth consists. There were those (for example, the followers of Epicurus) who rejected this view and assumed that all bodies in the universe--earth and stars--consist of the same material. Nevertheless, many educators and political leaders were afraid that denial of the exceptional status of the celestial bodies in physical science would make it more difficult to teach the belief in the existence of spiritual beings as distinct from material things; and since it was their general conviction that the belief in spiritual beings is a powerful instrument to bring about a desirable conduct among citizens, a physical theory that supported this belief seemed to be highly desirable.

Plato, in his philosophical treatise Laws, suggested that people in his ideal state who taught the "materialistic" doctrine about the constitution of sun and stairs would be jailed. We learn from this ancient example how scientific theories have served as instruments of indoctrination. Obviously, fitness to support the desirable conduct of citizens, or, briefly, to support moral behavior, has served

through the ages as a reason for acceptance of a theory. When the "scientific criteria" did not uniquely determine a theory, its fitness to support moral or political indoctrination became an important factor for its acceptance.

According to the passage, heated scientific debates frequently began because

a. a truly absurd theory had been proposed

b. a new theory jeopardized existing moral attitudes

c. scientist were unconcerned with politics and ethics

d. scientific theories had been adopted too hastily

e. a prevailing theory was incorrect but popular

Which of the following statements about Plato is supported by the passage?

a. He was the originator of the theory that the material of the celestial bodies differs from that of the earth

b. He was a proponent of the "materialistic" theory regarding the composition of the earth and stars

c. He was a lawmaker who punished offenders against society

d. He was an advocate of the censorship of ideas and theories

e. He was a political leader who founded and directed an ideal state

The main point of the passage is that scientific theories

a. are sometimes misinterpreted

b. ought not to be related to morality

c. have to be compatible with common sense

d. have changed through the ages

e. have often been used to support moral rules

Which of the following would be the most appropriate title for the passage?

a. The Case for Common Sense in Theory-Building

b. A Survey of "Unnatural" Scientific Theories

c. Factors in the Acceptance of Scientific Theories

d. the Moral Responsibility of Scientific Investigation

e. The Legacy of Greek Scientific Thought

The author's assertion that "We learn from the ancient example how scientific theories have served as instruments of indoctrination" (lines 48-50) is most weakened by the counterargument that

a. Plato was not describing a real event and his ideal state did not exist

b. ancient Greek ideas about matter were not really scientific theories

c. Plato believed that the "materialistic" doctrine would lead to immorality

d. the passage implied that scientific theories ought to be subordinate to the political needs of society

e. the passage implies that without indoctrination there would be no stable government